



Segmentation, Super-resolution and Fusion for Digital Mammogram Classification

By

TABAN FOUAD MAJEED

Department of Applied Computing
University of Buckingham
United Kingdom

A thesis submitted for the
Degree of Doctor of Philosophy in Computer Science to the
School of Science and Postgraduate Medicine in
The University of Buckingham

March 2016

ABSTRACT

Mammography is one of the most common and effective techniques used by radiologists for the early detection of breast cancer. Recently, computer-aided detection/diagnosis (CAD) has become a major research topic in medical imaging and has been widely applied in clinical situations. According to statistics, early detection of cancer can reduce the mortality rates by 30% to 70%, therefore detection and diagnosis in the early stage are very important. CAD systems are designed primarily to assist radiologists in detecting and classifying abnormalities in medical scan images, but the main challenges hindering their wider deployment is the difficulty in achieving accuracy rates that help improve radiologists' performance.

The detection and diagnosis of breast cancer face two main issues: the accuracy of the CAD system, and the radiologists' performance in reading and diagnosing mammograms. This thesis focused on the accuracy of CAD systems. In particular, we investigated two main steps of CAD systems; pre-processing (enhancement and segmentation), feature extraction and classification. Through this investigation, we make five main contributions to the field of automatic mammogram analysis.

In automated mammogram analysis, image segmentation techniques are employed in breast boundary or region-of-interest (ROI) extraction. In most Medio-Lateral Oblique (MLO) views of mammograms, the pectoral muscle represents a predominant density region and it is important to detect and segment out this muscle region during pre-processing because it could be bias to the detection of breast cancer. An important reason for the breast border extraction is that it will limit the search-zone for abnormalities in the region of the breast without undue influence from the background of the mammogram. Therefore, we propose a new scheme for breast border extraction, artifact removal and removal of annotations, which are found in the background of mammograms. This was achieved using an local adaptive threshold that creates a binary mask for the images, followed by the use of morphological operations. Furthermore, an adaptive algorithm is proposed to detect and remove the pectoral muscle automatically.

Feature extraction is another important step of any image-based pattern classification system. The performance of the corresponding classification depends very much on

how well the extracted features represent the object of interest. We investigated a range of different texture feature sets such as Local Binary Pattern Histogram (LBPH), Histogram of Oriented Gradients (HOG) descriptor, and Gray Level Co-occurrence Matrix (GLCM). We propose the use of multi-scale features based on wavelet and local binary patterns for mammogram classification. We extract histograms of LBP codes from the original image as well as the wavelet sub-bands. Extracted features are combined into a single feature set. Experimental results show that our proposed method of combining LBPH features obtained from the original image and with LBPH features obtained from the wavelet domain increase the classification accuracy (sensitivity and specificity) when compared with LBPH extracted from the original image.

The feature vector size could be large for some types of feature extraction schemes and they may contain redundant features that could have a negative effect on the performance of classification accuracy. Therefore, feature vector size reduction is needed to achieve higher accuracy as well as efficiency (processing and storage). We reduced the size of the features by applying principle component analysis (PCA) on the feature set and only chose a small number of eigen components to represent the features. Experimental results showed enhancement in the mammogram classification accuracy with a small set of features when compared with using original feature vector.

Then we investigated and propose the use of the feature and decision fusion in mammogram classification. In feature-level fusion, two or more extracted feature sets of the same mammogram are concatenated into a single larger fused feature vector to represent the mammogram. Whereas in decision-level fusion, the results of individual classifiers based on distinct features extracted from the same mammogram are combined into a single decision. In this case the final decision is made by majority voting among the results of individual classifiers.

Finally, we investigated the use of super resolution as a pre-processing step to enhance the mammograms prior to extracting features. From the preliminary experimental results we conclude that using enhanced mammograms have a positive effect on the performance of the system. Overall, our combination of proposals outperforms several existing schemes published in the literature.

*I dedicate my thesis to my Father,
Mother, and
To my husband*

ACKNOWLEDGMENT

Allah the most gracious and merciful: Who gave me the energy, health and courage to spend the time to complete this work.

Special thanks go to my supervisors, Dr Harin Sellahewa and Dr Naseer Al-Jawad, for their help, guidance and supervision.

I would like to thank my husband, Rasber, for his patience and endless support and my lovely angels, Shanya and Lanya. Many thanks go to my parents for their unconditional love and support all the way through. Also, thanks to all my friends for their support and encouragement.

I would like to thank Dr Sheeba Taneja, Consultant Radiologist at Milton Keynes Hospital, for her time and consultations.

Many thanks go to Dr Nadia Al-Hassan for her time and help with converting the original mammogram images to super resolved images.

ABBREVIATIONS

χ^2	chi-square test
ACR	American College of Radiology
AD	Architectural Distortions
ANN	Artificial Neural Networks
ASM	Angular Second Moment
BI-RADS	Breast Imaging Reporting and Data System
CAD	Computer Aided Detection/Diagnosis
CBIR	Content-Based Image Retrieval System
CC	Cranio-Caudal view
CLAHE	Contrast Limited Adaptive Histogram Equalization
CS	Compressive sensing
DDSM	Digital Database for Screening Mammography
DWT	Discrete Wavelet Transform
FN	False Negative
FP	False Positive
GLCM	Gray Level Co-occurrence Matrix
HDM	Hausdorff Distance Measure
HH	High-High (one of the wavelet sub-band)
HL	High-Low (one of the wavelet sub-band)
HOG	Histogram of Oriented Gradient
HRT	Hormone Replacement Therapy
ICA	Independent Component Analysis
IDM	Inverse Difference Moment
KNN	K-Nearest Neighbour
$LBP_{p,R}^{u2}$	Local Binary Patterns uniform pattern
LBP	Local Binary Pattern
LBPH	Local Binary Patterns Histogram
LCIS	Lobular Carcinoma in Situ
LH	Low-High (the wavelet sub-band representing horizontal/vertical features)

LL	Low-Low (the wavelet sub-band representing horizontal/vertical features)
LR	Low Resolution
Mini-MIAS	Mini-Mammographic Image Analysis Society
MLO	Medio-Lateral-Oblique view
MRF	Markov Random Field
MRI	Magnetic Resonance Image
PCA	Principle Component Analysis
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RP	Random Projection
SM	Spiculated Masses
SR	Super Resolution
SRG	Region Growing
ST-ROI	Surrounding Tissue Regions of Interest
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TN	True Negative
TP	True positive
WT	Wavelet Transform

CONTENTS

ABSTRACT.....	i
ACKNOWLEDGMENT.....	iv
ABBREVIATIONS	v
CONTENTS.....	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiv
DECLARATION	xviii
Chapter 1 Introduction	1
1.1 Challenges in a CAD System	3
1.2 Thesis Motivation.....	4
1.3 Thesis Aims and Contributions	4
1.4 Thesis Outline.....	5
1.5 List of Publications.....	6
Chapter 2 Background.....	8
2.1 Breast Cancer	8
2.1.1 What is Cancer?	8
2.1.2 What is Breast Cancer?	9
2.1.3 Breast Tumours	10
2.1.4 Risk Factors.....	11
2.2 Mammography	12
2.2.1 Objects (Lesions) in Mammogram Image.....	12
2.2.2 Mammograms Projection	14
2.2.3 Breast Image Reporting.....	14
2.2.4 Challenges in Reading Mammogram.....	15
2.3 CAD Systems	16

2.3.1	What is a CAD System?.....	16
2.3.2	CAD Architecture	17
2.3.3	Evaluation of CAD System.....	18
2.4	Mammogram Image Databases	20
2.4.1	Mini-MIAS Database	21
2.4.2	DDSM Database.....	22
Chapter 3 Literature review		23
3.1	General CAD System	23
3.2	Pre-processing (Segmentation).....	25
3.3	Feature Extraction and Classification.....	30
3.4	Summary	36
Chapter 4 Breast Border Extraction and Pectoral Muscle Removal in MLO Mammogram.....		38
4.1	Breast Region Segmentation	38
4.2	Breast Border Extraction	40
4.3	Pectoral Muscle Removal.....	42
4.4	Experiment Setup	45
4.4.1	Dataset and Strategy.....	46
4.4.2	Gray level Co-occurrence Matrix Features	46
4.4.3	Classification Methods.....	51
4.5	Results and Discussion	52
4.6	Conclusions	58
Chapter 5 Texture-based Features for Mammogram Classification		60
5.1	Background	61
5.1.1	Discrete Wavelet Transform (DWT)	61
5.1.2	Local Binary Patterns (LBP) Features	63
5.1.3	Multi-scale LBP	66

5.1.4	Histogram of Oriented Gradients (HOG) descriptor.....	66
5.1.5	Gray level co-occurrence matrix (GLCM).....	67
5.2	Proposed Feature Extraction Methods	68
5.2.1	Multi-Scale LBP	68
5.2.2	GLCM	69
5.2.3	HOG Descriptor	70
5.3	Experimental Setup	71
5.3.1	Databases.....	71
5.3.2	Evaluation Protocol.....	72
5.3.3	Classification Methods.....	73
5.4	Experimental results and discussion.....	73
5.4.1	Multi-Scale LBP	73
5.4.2	GLCM	81
5.4.3	HOG	82
5.5	Results Summary.....	83
5.6	Dimension Reduction	85
5.7	Dimension Reduction on Proposed Schemes	88
5.7.1	PCA over HOG	88
5.7.2	PCA over LBP	89
5.7.3	Experimental Setup and Results	89
5.8	Conclusions	94
Chapter 6 Fusion for Mammogram Classification.....		97
6.1	Fusion Techniques.....	98
6.1.1	Feature-Level Fusion	98
6.1.2	Decision-Level Fusion	98
6.2	Feature and Decision Fusion for Mammogram Classification.....	100

6.2.1	Feature-Level Fusion for Mammogram Classification	100
6.2.2	Proposed Decision-Level Fusion methods	105
6.3	Statistical evaluation of Experiment Results	107
6.3.1	Significance between the observed and expected model	107
6.3.2	SVM Score Analysis	112
6.4	Comparison with Existing Works	114
6.5	Conclusions	121
Chapter 7 Mammogram Enhancement using Super resolution.....		124
7.1	Super Resolution Technique.....	125
7.2	Experimental Result	127
7.2.1	Original LBP	128
7.2.2	Multi Sub-bands LBP.....	129
7.2.3	Original LBP with Multi Sub-bands LBP	132
7.2.4	HOG Descriptor	135
7.2.5	PCA-HOG and PCA-LBP.....	136
7.2.6	GLCM	137
7.2.7	Decision level fusion.....	138
7.3	Conclusions	141
Chapter 8 Conclusions and Future Work		143
8.1	Limitations and Future Works.....	147
REFERENCES.....		149
APPENDIX A: Some examples of SR-Mini-MIAS database		162
APPENDIX B: Super resolution Mini-MIAS database		164
APPENDIX C: Super resolution DDSM database		171
APPENDIX D: Feature Fusion		179

LIST OF FIGURES

Figure 1.1 : Cranial-Caudal (CC) view and Medio-Lateral-Oblique View (MLO) (Molloy, 1997)	2
Figure 2.1: The part of female breast: image from www.breastcancer.org	9
Figure 2.2 Shapes of mass (adapted from (Yacoub, et al., 2006)).....	13
Figure 2.3: Mammogram Projections	14
Figure 2.4: CAD architecture	17
Figure 2.5: Examples of different tissue types available in the Mini-MIAS database	21
Figure 2.6: Sample mass ROIs first row and normal ROIs second row	22
Figure 4.1: Types of noises example observed at a mammogram mdb002 in Mini-MIAS database (Suckling, et al., 1994)	39
Figure 4.2: Example images of the Mini-MIAS database with different types of noises and background artifacts	39
Figure 4.3: Breast border extraction process.....	41
Figure 4.4: (a) Original image, (b) Binarised image, (c) Removed artifacts, (d) Extracted breast region.....	42
Figure 4.5: Pectoral muscle removal process.....	44
Figure 4.6: Output of the Pectoral muscle removal algorithm.....	45
Figure 4.7: Example of extracting GLCM matrix from a part of image.....	47
Figure 4.8: An example of SVM model.....	52
Figure 4.9: Label removal and Border extraction sample.....	53
Figure 4.10: Radiologist Decisions for muscle removal.....	54
Figure 4.11: Classification accuracy by KNN classifier, (a) $d=1$ and angle=0 degree, (b) $d=1$ and angle= 45 degree, (c) $d=1$ and angle= 90 degree and (d) $d=1$ and angle=135 degree	58
Figure 4.12: Classification accuracy by SVM classifier, (a) $d=1$ and angle=0 degree, (b) $d=1$ and angle= 45 degree, (c) $d=1$ and angle= 90 degree and (d) $d=1$ and angle=135 degree	58

Figure 5.1: Wavelet Decompositions	62
Figure 5.2: Three Level Pyramid DWT decomposition of mammograms	63
Figure 5.3: An example of LBP operator	64
Figure 5.4: Sub block histogram concatenating (Ahonen, et al., 2004).....	65
Figure 5.5: Circular neighbourhoods for LBP (a, (b), and (c)	66
Figure 5.6 : Results summary of Mini-MIAS database accuracy	83
Figure 5.7 : Results summary of DDSM database accuracy.....	84
Figure 5.8: Largest and smallest Principle components.....	88
Figure 5.9: General frame work of PCA-HOG	89
Figure 6.1: General frame work of feature level fusion strategy	99
Figure 6.2: General frame work of the decision level fusion strategy	100
Figure 6.3 : Proposed HOG and LBPH features fusion diagram.....	101
Figure 6.4 : Proposed PCA-HOG and PCA-LBPH features fusion diagram.....	104
Figure 6.5: Examples of Correctly and Incorrectly classified mammograms.....	123
Figure 7.1: An example of Mini-MIAS mammogram: (a) Normal LR-Mini-MIAS mammogram, (b) Normal SR-Mini-MIAS mammogram, (c) Abnormal LR-Mini- MIAS mammogram, and (d) Abnormal SR-Mini-MIAS mammogram	127
Figure 7.2: Results of Mini-MIAS Database based on LBPH scheme obtained from original mammogram	128
Figure 7.3: Results of DDSM Database based on LBPH scheme obtained from original mammogram	128
Figure 7.4: Multi sub-bands results of leave-one-out strategy, (a) Mini-MIAS, (b) DDSM	130
Figure 7.5: Multi sub-bands results of 70% training and 30% testing strategy, (a) Mini- MIAS, (b) DDSM	130
Figure 7.6: Multi sub-bands results of 50% training and 50% testing strategy, (a) Mini- MIAS, (b) DDSM	131

Figure 7.7: Multi sub-bands results of 10% training and 90% testing strategy, (a) Mini-MIAS, (b) DDSM	131
Figure 7.8: Original LBP and multi sub-bands LBP results of leave-one-out strategy, (a) Mini-MIAS, (b) DDSM	133
Figure 7.9: Original LBP and multi sub-bands LBP results of 70% / 30% strategy, (a) Mini-MIAS, (b) DDSM	133
Figure 7.10: Original LBP and multi sub-bands LBP results of 50% / 50% strategy, (a) Mini-MIAS, (b) DDSM	134
Figure 7.11: Original LBP and multi sub-bands LBP results of 10% / 90% strategy, (a) Mini-MIAS, (b) DDSM	134
Figure 7.12: Results of HOG features obtained from Mini-MIAS Database	135
Figure 7.13: Results of HOG features obtained from DDSM Database	135
Figure 7.14: Results based on PCA-LBP features obtained from (a) Mini-MIAS, (b) DDSM Database.....	136
Figure 7.15: Results based on PCA-HOG features obtained from (a) Mini-MIAS, (b) DDSM Database.....	137
Figure 7.16: Results of GLCM features obtained from Mini-MIAS database	138
Figure 7.17: Results of using four strategies based on GLCM features obtained from DDSM database	138
Figure 7.18: Decision fusion results of Leave-one-out strategy, (a) Mini-MIAS, (b) DDSM	139
Figure 7.19: Decision fusion results of 70% training and 30% testing strategy, (a) Mini-MIAS, (b) DDSM	139
Figure 7.20: Decision fusion results of 50% training and 50% testing strategy, (a) Mini-MIAS, (b) DDSM	140
Figure 7.21: Decision fusion results of 10% training and 90% testing strategy, (a) Mini-MIAS, (b) DDSM	140

LIST OF TABLES

Table 2.1: CAD system evaluation terminology.....	19
Table 2.2: CAD measure example	20
Table 3.1: Pre-processing (segmentation) techniques summary.....	30
Table 3.2: Feature extraction and classification summary.....	35
Table 4.1: Classification result, $d=1$, angle=0 degree.....	54
Table 4.2: Classification result, $d=1$, angle=45 degree.....	55
Table 4.3: Classification result, $d=1$, angle=90 degree.....	55
Table 4.4: Classification result, $d=1$, angle=135 degree.....	56
Table 5.1: Results of leave-one-out strategy based on LBPH features obtained from single wavelet sub-band	74
Table 5.2: Results of 70%/30% training/testing strategy based on LBPH features obtained from single wavelet sub-band.....	74
Table 5.3: Results of 50%/50% training/testing strategy based on LBPH features obtained from single wavelet sub-band.....	75
Table 5.4: Results of 10%/90% training/testing strategy based on LBPH features obtained from single wavelet sub-band.....	75
Table 5.5: Results of leave-one-out strategy based on LBPH features obtained from multi wavelet sub- bands.....	76
Table 5.6: Results of 70%/30% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands	76
Table 5.7: Results of 50%/50% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands	77
Table 5.8: Results of 10%/90% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands	77
Table 5.9: Results of different strategy based on LBPH features obtained from original mammogram images	78

Table 5.10: Results of leave-one-out tests strategy based on LBPH features obtained from original mammogram and wavelet sub-bands.....	78
Table 5.11: Results of 70%/30% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands.....	79
Table 5.12: Results of 50%/50% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands.....	79
Table 5.13: Results of 10%/90% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands.....	80
Table 5.14: Results of leave-one-out tests strategy based on 14 GLCM features obtained from the original mammogram.....	81
Table 5.15: Results of 70%/30% training/testing strategy based on 14 GLCM features obtained from the original mammogram.....	81
Table 5.16: Results of 50%/50% training/testing strategy based on 14 GLCM features obtained from the original mammogram.....	82
Table 5.17: Results of 10%/90% training/ testing strategy based on 14 GLCM features obtained from the original mammogram.....	82
Table 5.18: Results of using four strategies based on HOG features obtained from original mammogram.....	82
Table 5.19: Feature vector size	86
Table 5.20: Results of leave-one-out strategy based on applying PCA over HOG features with selecting various PCA components.....	89
Table 5.21: Results of 70%/30% training/testing strategy based on applying PCA over HOG features with selecting various PCA components	90
Table 5.22: Results of 50%/50% training/testing strategy based on applying PCA over HOG features with selecting various PCA components	90
Table 5.23: Results of 10%/90% training/testing strategy based on applying PCA over HOG features with selecting various PCA components	91
Table 5.24: Results of leave-one-out strategy based on applying PCA over LBPH features with selecting various PCA components.....	92

Table 5.25: Results of 70%/30% training/testing strategy based on applying PCA over LBP features with selecting various PCA components.....	92
Table 5.26: Results of 50%/50% training/testing strategy based on applying PCA over LBP features with selecting various PCA components.....	93
Table 5.27: Results of 10%/90% training/testing strategy based on applying PCA over LBP features with selecting various PCA components.....	93
Table 6.1: Results of leave-one-out tests strategy based on fusing HOG and LBPH features	101
Table 6.2 : Results of 70%/30% training/ testing strategy based on fusing HOG and LBPH features	102
Table 6.3 : Results of 50%/50% training/testing strategy based on fusing HOG and LBPH features	102
Table 6.4 : Results of 10%/90% training/testing strategy based on fusing HOG and LBPH features	103
Table 6.5 : Results of leave-one-out tests strategy based on fusing PCA-HOG and PCA-LBPH features	104
Table 6.6 : Results of 70%/30% training/testing strategy based on fusing PCA-HOG and PCA-LBPH features	104
Table 6.7 : Results of leave-one-out tests strategy based on decision level fusion of using different kind of f features	106
Table 6.8 : Results of 70%/30% training/testing strategy based on decision level fusion of using different kind of features	106
Table 6.9: Contingency table shows the frequencies of 1s and 0s for Method1 and Method2	108
Table 6.10: Chi-square (x²) test for mammogram image classification on Mini-MIAS database	110
Table 6.11: Chi-square (x²) test for mammogram image classification on DDSM database	111
Table 6.12: SVM score analysis for Mini-MIAS database	113
Table 6.13: SVM score analysis for DDSM database.....	114

Table 6.14: Comparison between proposed method and some related work (Mini-MIAS database).....	115
Table 6.15: Comparison between proposed method and some related work (DDSM database).....	117

DECLARATION

I, hereby declare that presented work has not previously been submitted towards any qualification, degree or diploma at the University of Buckingham or any other university.

I also declare that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Taban Fouad Majeed

Chapter 1

Introduction

Breast cancer is a major global health problem in the world. It is the main cause of mortality amongst women. Excluding skin cancers, breast cancer is the most commonly diagnosed cancer amongst women in the USA. Also, it comes as the second leading cause of cancer-related death (Kwan¹, et al., 2009). It affects women over the age of 15 (usually between 35-55 years) as well as men over the age of 40 (Senthilkumar & Umamaheswari, 2011). Reports from 2007 have shown an estimated 178,480 newly-diagnosed cases of breast cancer and 40,460 deaths from the disease amongst women (Tang, et al., 2009). Moreover, reports from the World Health Organization's (WHO) International Agency for Research on Cancer in Lyon-France, shows that more than 150,000 women around the world die from breast cancer each year (Yu & Guan, 2000). Given the high mortality rates associated with breast cancer, early detection remains the key to ameliorating its prognosis (Xue, et al., 2010) (Deserno, et al., 2011) (Khuzi, et al., 2009). The death rate in the UK shows that for every 100000 females there are around 36 death cases caused by breast cancer, while the rate is less than 1 for every 100000 males (Cancer Research, UK, 2014).

Mammography is known as one of the effective tools for earlier breast cancer detection. It is a screening tool used to examine the human breast by using low-dose amplitude X-rays. Here, asymptomatic women with no clinical suspicion for breast cancer are screened using screening mammography and this has reduced the mortality rate by 30-70% (Linguraru, et al., 2006). Two imaging projections of each breast are routinely obtained as part of the human test, Cranio-Caudal (CC) and Medio-Lateral-Oblique (MLO) views, as shown in Figure 1.1.

When reading a mammogram, in the majority of cases, a specialist radiologist would identify a radiographic abnormality; however, there is a chance of missing the same abnormality under different circumstances. So, the probability of false negatives is high. Retrospective studies (Christoyianni, et al., 2002) (Winsberg, et al., 1967) (Marx, et al., 2004) have shown that 10 to 30% of the cancers are undetected due to radiologist fatigue, the complex structure of the breast tissue, and the subtlety of the cancer. It was

hoped that with the aid of Computer Aided Detection (CAD) schemes, this lack of consistency of the human observer may be overcome when used as a reminder or “second opinion” (Masala, 2006) (Sampat, et al., 2008). Results have shown that the performance of radiologists can be improved by providing them with the results of a CAD system. Radiologist's performance can be improved by providing them with the results of a CAD system.

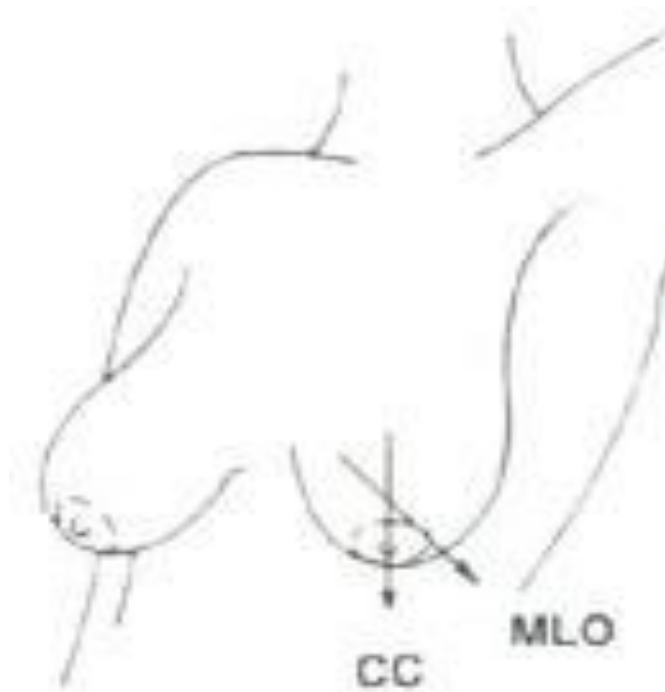


Figure 1.1 : Cranial-Caudal (CC) view and Medio-Lateral-Oblique View (MLO) (Molloy, 1997)

Nowadays, Computer-aided detection/diagnosis (CAD) is an integral part of the process adopted by radiologists and Oncology specialists in detecting breast cancer from mammograms produced at most screening centres and hospitals. In fact, over the last few years CAD has been developed and incorporated in the detection and differential diagnosis of a variety of abnormalities in medical images obtained from different scanning schemes (Doi, 2007).

Therefore, it was felt that a high performance CAD system is needed to be designed in order to help radiologists to detect cases of breast cancer early, hence, reducing, or even eliminating, the rate of mortality amongst women with the disease.

In this thesis we will investigate a method to breast border extraction and pectoral muscle removal. Also we classify the mammogram to normal and abnormal by

extracting textural feature and fuse the features in two different level (feature level fusion and classification level fusion). Finally, we will investigate the use of super-resolution methods to enhance the mammogram classification.

1.1 Challenges in a CAD System

The large set of images produced by screening mammogram tests constitutes a huge workload for the relatively few radiologists that need to interpret these mammogram images. Thus, the computerized mammographic analysis will have a great impact in terms of easing the workload and assisting in the detection of the breast cancer.

In CAD systems, detecting an abnormality is difficult because of the characteristics and structure of breast abnormalities. Sometimes mass and micro-calcifications are superimposed and hidden in the dense tissue, therefore this makes the segmentation of a correct region of interest (ROI) difficult. A wide range of features of the abnormality (e.g. shape, margin and density) and their low visibility within the surrounding tissue makes the computer aided detection and diagnosis of breast abnormalities a challenge.

A typical CAD system includes a number of steps: pre-processing and segmentation, feature extraction, feature reduction (selection), and classification. Each step has its challenges and the outcome of one affecting the result of the next. Before a digitized mammogram is analyzed by a computer, it must be segmented into its representative anatomical regions. Accurate segmentation of the breast region in mammograms is a key pre-processing step in the analysis of mammograms. This step is essential for confining the relevant search regions that may contain the suspect abnormalities and reduce excessive influence from the irrelevant background regions of the mammogram.

Examples of background information that must be segmented and excluded from the mammogram region are types of noise observed in mammograms that include high intensity rectangular label, low intensity label and tape artifacts (Tzikopoulos, et al., 2011).

The pectoral muscle should always appear as a high-intensity, triangular region across the upper posterior margin of the image on a proper MLO view. We deal with this kind of challenge in details in chapter four of this thesis by proposing breast border extraction and pectoral muscle suppression as a part of the pre-processing step. Automatically extracting and selecting appropriate features that will give the best classification accuracy results is a difficult task. Furthermore, the choice of a classifier

has a great influence on the final result. Again we deal with these kinds of challenges in detail by investigating several features extraction techniques, feature fusion, feature reduction (selection), and classification in chapters five and six of this thesis. Finally, in chapter seven we will investigate the use of super-resolution to enhance the mammogram prior to feature extraction.

1.2 Thesis Motivation

Since the cause of breast cancer is still not certain, prevention against breast cancer remains a challenge. However, there is a chance of complete recovery if the efficient diagnosis of breast cancer was made at an early stage. In other words, early detection of breast cancer can play a key role in reducing the associated morbidity and mortality rates. Advances have been made in the diagnosis of breast cancer at earlier stages of development using radiographic breast imaging and screening programs; however, figures show that 10-30% of biopsy-proven malignant cases are missed due to various factors and reasons such as technical problems in the imaging procedure, abnormalities that are not apparent, and abnormalities that are misread (Shinde, 2003).

As mentioned above, a high percentage of breast cancer cases can be missed by conventional screening mammography. Therefore, this fact motivates us to create and design Computer-aided detection/diagnosis (CAD) systems, which use advanced image analysis techniques to detect abnormalities in mammograms. Separating mammogram images to the normal and abnormal case is defined as the first step towards detecting abnormality or to help radiologists in making a diagnosis.

1.3 Thesis Aims and Contributions

A high performance CAD system is needed to be designed in order to help radiologists to detect cases of breast cancer in an early stage, thereby, reducing, or even eliminating, the rate of mortality amongst women with the disease.

Our primary objective is to investigate and develop reliable algorithms designed to automate the computer- based aspects of the process of detecting and classifying abnormalities in digital mammograms. In particular we aim to eventually provide a Computer-Aided Detection/Diagnosis (CAD) system to improve existing tools and help facilitate comprehensive screening systems.

In order to satisfy our aims expressed above, the main contributions of this thesis can be stated as follows:

- **Developed and proposed** a scheme to remove irrelevant parts of the mammograms to prepare the image for feature extraction. This resulted in obtaining a higher mammogram classification accuracy when compared to using all mammograms parts. Irrelevant parts are breast region background and muscles of the breast, especially in MLO view images which causes a negative effect on the classification accuracy of the CAD system.
- Different types of texture feature extraction schemes such as Local Binary Pattern (LBP), wavelet based multiscale LBP, Gray Level Co-occurrence Matrix (GLCM), and Histogram Oriented Gradient (HOG), have been investigated. We propose multi-scale features based on wavelet and local binary patterns for mammogram classification. Furthermore, two different classifiers, namely K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) are discussed and investigated.
- Propose to use PCA to reduce the dimension of very large feature sets (e.g. HOG features, multiscale LBPH) and demonstrated that a very small feature vector could be used without compromising classification accuracy. The small feature sets improve efficiency (processing and storage) of the system.
- Propose the use of the feature and decision fusion to obtain better mammogram classification accuracy and overcome the limitation of using single feature type mentioned above. We show that decision level fusion can lead to significant improvement in classification accuracy, similar to a number of radiologists examining the same mammogram to arrive at a diagnosis.
 - Propose the use of super resolution to enhance mammograms prior to feature extraction. Experimental results show an improvement in classification accuracy. However, further research is required to establish a deeper understanding of the effects of super-resolving mammograms on the final classification.

1.4 Thesis Outline

The rest of the thesis is organised as follows:

- **Chapter two:** A theoretical background about the main focuses subjects are given in details. Focused subjects are breast cancer, mammography, and CAD system. For each subject, important terminology and concepts are explained such as in the breast cancer field, the risk factors are given, in mammography field, mammography projection and challenges of reading mammogram images

explained, while in the CAD system field, the architecture of any CAD system given as well as evaluation measurements of any CAD system are discussed. At the end of the chapter properties of used databases are given.

- **Chapter three:** provides literature survey about the CAD system, pre-processing step, and feature extraction methods to classify mammogram images.
- **Chapter four:** proposes a method for mammogram border extraction and pectoral muscle removal. The experimental results show that using segmented mammogram images by our proposed method has better classification accuracy than using original mammograms.
- **Chapter five:** a number of texture-based features are investigated for mammogram classification. We propose the use of multiscale LBPH features in combination with existing features to represent mammograms. Results of experiments on two databases are discussed here. Also we present the use of PCA to reduce the number of features while keeping the classification accuracy at the same level or even increase it when compared with using original feature sets.
- **Chapter six:** proposes the use of the feature and decision fusion to increase mammogram classification accuracy.
- **Chapter seven:** proposes the use of super resolution to enhance mammograms prior to feature extraction. Several experiments will be conducted to evaluate the effects of super resolution on mammogram classification.
- **Chapter eight:** ends the thesis by presenting our overall conclusions and direction of future research work.

1.5 List of Publications

Peer Reviewed Paper:

1. Taban F. Majeed, Naseer Al-Jawad and Harin Sellahewa, “Breast Border Extraction and Pectoral Muscle Removal in MLO Mammogram Images”, Fifth Computer Science and Electronic Engineering Conference (CEEC 2013), September 17-18, 2013, University of Essex, Colchester, United Kingdom.

Posters:

1. Taban F. Majeed, Naseer Al-Jawad and Harin Sellahewa, “Breast Border Extraction and Pectoral Muscle Removal in MLO Mammogram Images”, in BioTrinity 2013 - European Biopartnering and Investment Conference - May 14-16, 2013, Newbury Berkshire, United Kingdom.
2. Taban F. Majeed, Naseer Al-Jawad and Harin Sellahewa, “Breast Border Extraction and Pectoral Muscle Removal in MLO Mammogram Images”, in 9th London Hopper poster competition, 23 May 2013, London, United Kingdom.

Chapter 2

Background

This chapter gives a general background to cancer and includes definitions of cancer especially breast cancer, types of cancer appears in the breast, and risk factors of the breast cancer. Mammography and its use in detecting the breast abnormality are explained in Section 2.2. A brief description of CAD systems, their advantages for detecting breast abnormality, system architecture, and system evaluation are presented in Section 2.3. Finally, the mammogram databases used in this thesis will be described in Section 2.4. Readers familiar with the general background on breast cancer and mammography may skip Sections 2.1 and 2.2.

2.1 Breast Cancer

This section presents important background information on breast cancer.

2.1.1 What is Cancer?

The human body is composed of millions of cells. Normally, body cells grow, multiply and then die in a programmed and controlled way. The rate at which body cells divide and grow varies with age. In infancy and childhood, normal cells divide faster to allow for growth. In adult life, some cells differentiate and specialize and hence stop dividing, while others divide only to replace cells lost either through injury or after cell death. The whole processes of cell growth and division is under close monitoring by different mechanisms in the body to keep the processes well controlled.

A group of conditions where the human body's cells begin to grow and divide in an uncontrolled fashion is called cancer. When cancer arises from epithelial cells it is called 'carcinoma'. Six biological capabilities characterise cancer cells and they constitute what is called 'hallmarks of cancer'. These biological capabilities are obtained during the multistep development of human tumours and include self-sufficiency in growth signals, evading growth suppressors, evasion of apoptosis (programmed cell death), limitless, replicative potential, the induction of angiogenesis, and tissue invasion and metastasis, cancer cells seize nutrients from normal cells, and

then encroach on the surrounding tissues. Cancer cells can even detach from tumour and metastasis to other parts of the human body via the lymphatic or vascular systems (i.e. spreading of cancer from one part of the body to another). Genome instability is pivotal in the development and maintenance of these hallmarks (Samulski, 2006) (Sample, 2003) (Hanahan & Weinberg, 2011) (Che Kuo, et al., 2014).

2.1.2 What is Breast Cancer?

Breast cancer is the most common cancers in the UK. Each year, approximately 50,000 women are affected by breast cancer in the UK. Amongst these, 80% are over 50; however, younger women, and in rare cases men, can also get the disease (UK.GOV, 2010). In the United States the breast cancer has become the second leading cause of cancer death in women. The World Health Organization's statistics show that around 519,000 women worldwide will die of breast cancer (Elshinawy, 2010).

The female breast is composed of lobules (glands that produce milk), ducts (the milk carrier from the lobules to the nipple), and stroma (adipose and connective tissue surrounding the ducts and lobules with blood and lymphatic vessels). Figure 2.1 show the anatomy of female breast (da Fonseca, February 2013) (NIH, September, 2009).

Breast profile	
A	Ducts
B	Lobules
C	Dilated section of duct
D	Nipple
E	Fat
F	Pectorals major muscle
G	Chest wall/rib cage
Enlargement	
A	Normal duct cells
B	Basement membrane
C	Lumen (centre of duct)

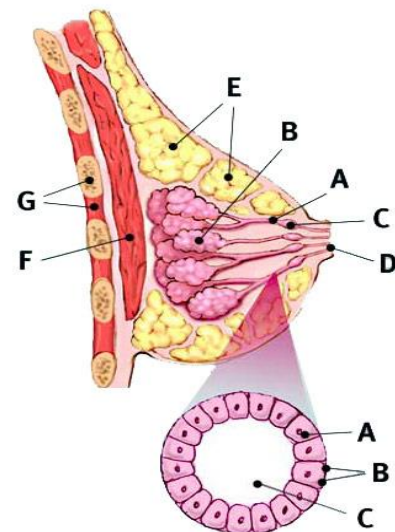


Figure 2.1: The part of female breast: image from www.breastcancer.org

Many cases of breast cancer arise from cells that line the breast ducts (ductal cancers), but cells that line the lobules are also the source of other cancers (lobular cancers). Breast cancer has adverse effects on mammary gland and cancer cells can metastasize (i.e. break away from the breast tumour and disperse to other parts of the body via the lymphatic system. Therefore, the lymphatic drainage of the breasts is very relevant to oncology (Society, 2009).

2.1.3 Breast Tumours

Breast tumours can be classified into three types: benign breast tumours, in situ breast cancers, and invasive breast cancers (Samulski, 2006) (Sample, 2003).

2.1.3.1 Benign Breast Tumours

This type of breast neoplasm forms the majority of breast tumours detected by mammography. As the name implies, these are noncancerous growths and lack the capability to invade surrounding tissues or spread to other organs. Most benign tumour cases can be left alone if they show no symptoms and create no complications. The examinations or imaging scans need to ensure that the tumour is not growing. In some cases it is difficult to distinguish certain benign masses from malignant lesions with mammography scans.

2.1.3.2 In Situ Cancer

In situ breast cancer (non-invasive breast cancer) is where the cancer cells have not invaded the basal membrane and are fully contained in the lobule or the duct. It is at a stage where it has not spread to the surrounding tissues in the breast or other parts of the body. However, it has the potential to develop into a more serious invasive cancer.

Depending on whether the carcinomatous cells arise from the ducts or the lobules, two types of non-invasive breast cancer are available: (i) ductal carcinoma in situ (DCIS) which is named as stage 0 cancer that has not spread beyond the ductal system (i.e. to the lymph nodes or other parts of the human body). The chance of overcoming this kind of cancer can be increased by taking the proper treatment, (ii) lobular carcinoma in situ (LCIS), the LCIS type of cancer is more difficult to detect using the mammography compared with DCIS cases. LCIS cases are usually discovered when the patient is being investigated and biopsied for another kind of abnormality. Therefore for early detection of breast cancer, the recommendation by American Cancer Society includes

a clinical examination every three years for women aged between 20 to 39 years, and at the age 40, the annual mammography screening should start. Another optional recommendation of monthly breast self-examination starts from 20 years old is made to the American Cancer Society.

2.1.3.3 Invasive Cancer

Cancer is said to be invasive when cancer cells have broken through the basal membrane and spread into the surrounding tissue. Here the chance of metastasis increases considerably. The American Cancer Society shows that age is an important factor in developing breast cancer among women, so women under age 45 have more chance of contracting invasive breast cancer diseases. Unfortunately, **two thirds** of invasive breast cancer cases are only diagnosed at the age of 55 or more (Samulski, 2006).

Other risk factors include race, geography, socioeconomic status, family history, genetics, exposure to radiation, obesity, and alcohol intake (see Section 2.1.4 for risk factors). The most common form of breast cancer is invasive ductal breast cancer. There are other types of breast cancer which are classed as special types such as invasive lobular breast cancer and some rare types of breast cancer. Invasive ductal breast cancer accounts for about 80% of all cases of breast cancer. The invasive ductal breast cancer has a number of possible symptoms, for example change of the breast size and shape, lump or thickening of the breast tissue or skin puckering or dimpling of the skin, swelling under the arm, discharge from the nipple, and constant pain in the breast or armpit. Less commonly, a type of rash involving the nipple.

It is worth mentioning that the stage of breast cancer at the time of detection dictates the success of any subsequent treatment. The stage of a tumour often takes into account its size, depth of penetration whether there is any invasion to adjacent organs, whether there is metastasis to regional lymph nodes, and if so how many lymph nodes, and whether it has spread to distant organs .

2.1.4 Risk Factors

The exact causes of breast cancer are still not known; however, there are some factors that increase the risk of developing the disease. These include early menarche or late menopause, family history of breast cancer, taking combined hormone replacement therapy (HRT) containing oestrogen and progesterone, the risk of breast cancer is

strongly related to age, certain dietary factors and some lifestyle factors including excessive consumption of alcohol over many years, excessive smoking over many years, physical inactivity and being overweight (increased levels of fat tissue-derived oestrogen). Women with dense breast tissue (when the breast is mostly composed of glandular and connective tissue with scanty amount of fatty tissue) manifested in mammograms as a larger area of dense tissue for the breast if it compares with the mammogram of the women of the same age are at higher risk of breast cancer (Society, 2009).

2.2 Mammography

Many different techniques for imaging have been produced such as (X-ray, Magnetic Resonance Image (MRI), and ultrasound). However, mammography is the technique that is used most widely to detect breast cancer at early stages. Digital mammography represents an advanced form of x-ray mammography. Here, a special electronic x-ray detector is used to capture the breast image and the image is converted into a digital picture which is then stored on a computer. The magnification, orientation, brightness and contrast of the mammogram image may be changed after the exam is finished to enhance the clarity of the digital mammogram images and to help the radiologist to see certain zones clearly. The advantages of digital mammography over traditional film mammography are (i) faster image acquisition, (ii) examination time is shorter, (iii) easier processes of image storage, (iv) low cost when compared to other techniques such MRI, and (v) transmission to other health professionals. In addition, it facilitates computer processing analysis of breast images which results in greater accuracy in breast cancer detection (Hashimoto, 2008) (The Independent UK Panel on Breast Cancer, October 2012).

2.2.1 Objects (Lesions) in Mammogram Image

There are certain characteristic lesions that characterize breast cancer; these include microcalcifications, masses, and architectural distortions. Bilateral asymmetry may also hint towards breast cancer.

Microcalcifications are tiny lesions (typically 0.05 - 1 mm) on mammograms. Their minute sizes make them relatively difficult to detect. Beside their minute and variable sizes, they are bright, of different shapes and distribution, and in some cases of low contrast making them even harder to spot. Furthermore, their proximity to the

surrounding dense tissues adds to the difficulty in detection because of considerable overlapping between normal and suspicious tissues. To further complicate the picture, some anatomic structures such as fibrous strands, breast margins or hypertrophied lobules mimic microcalcifications in their mammographic appearance. Accurate identification of microcalcifications is paramount for an early detection of the majority of breast cancer cases, because their presence is strongly correlated with breast cancer, particularly when they appear in clusters (Bozek, et al., september, 2008) (de Cruz, July, 2011).

The morphology of the microcalcifications is an important indicator of the nature of the underlying pathology. Generally larger, round and oval shaped calcifications with uniform size are in favour of a benign lesion; whereas, calcifications that are smaller, irregular, heterogeneous in size and morphology and of arborizing nature are in favour of being malignant (de Cruz, July, 2011).

On mammograms, masses appear as dense regions in the breast that are of different sizes and properties. The categorization of masses is based on their shape, density, and margins. Regarding the shape, masses could be round, oval, lobular, or irregular as shown in Figure 2.2. The margins can be described as circumscribed (well-defined with clearly demarcated borders), or obscured (margins are not clear because of overlapping by adjacent tissue). With regard to density, masses could be either of high density, low density, equal density, or fat containing. This categorization helps radiologists in the precise description of masses found in mammograms and in commenting on masses as either benign or potentially malignant (Sample, 2003) (Rangayyan, 2005).

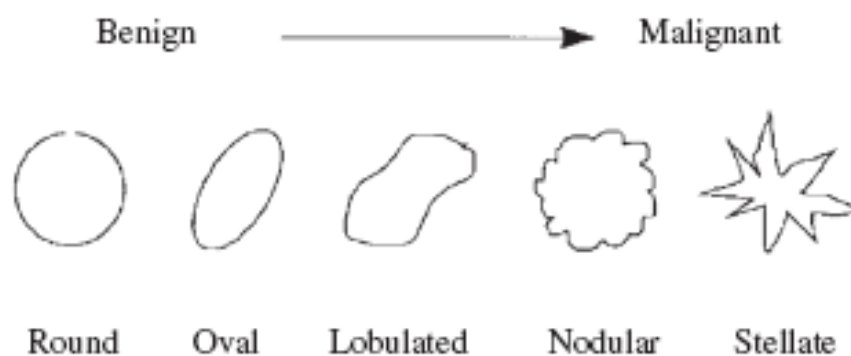


Figure 2.2 Shapes of mass (adapted from (Yacoub, et al., 2006))

Architectural distortions on a mammogram basically refer to a disruption in the normal radiating or arbitrary pattern of the parenchyma. With no visible centre or mass, they are very variable and hence very difficult to detect (de Cruz, July, 2011).

2.2.2 Mammograms Projection

The mammograms have two important projections of the breast:

- a) Medio-Lateral Oblique side view: this facilitates making almost the whole breast visible commonly including lymph nodes which taken at an angle. In this view, the pectoral muscle will appear in the upper part of the mammogram images.
- b) Cranio-Caudal view: CC view is one of the two standard projections in a screening mammography. The CC view is taken from top to bottom. It must show the medial part as well the external lateral portion of the breast as much as possible (Varjonen, et al., 2007). Figure 2.3 illustrates the two most common types of breast projections.

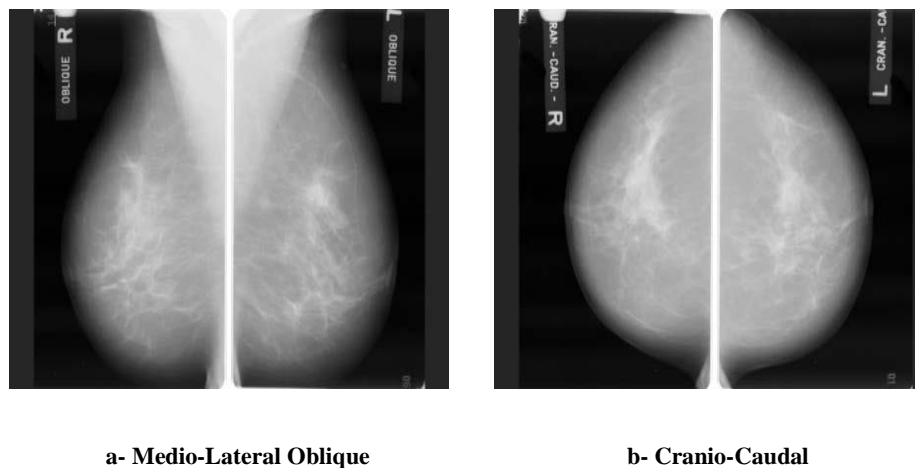


Figure 2.3: Mammogram Projections

2.2.3 Breast Image Reporting

BI-RADS are an acronym for Breast Imaging- Reporting and Data System. It was developed by the American College of Radiology (ACR). The system is designed to standardize reporting. Each category provides the overall assessment related to the findings and the necessary follow up. The radiologist assigns a BI-RADS category 0 to 6 as the final imaging result (Eberl, et al., 2005) (Hashimoto, 2008)

- Category 0 Incomplete Assessment: this type needs more imaging evaluation or prior mammogram images for comparison.
- Category 1 Negative: Nothing appears to comment on.

- Category 2 (Benign Finding): this type is negative mammogram images. Although in some cases, the radiologist may find something and describe it as calcified fibro adenomas, or fat-containing lesions such as (oil cysts)...etc. The mammogram did not show any evidence of malignancy.
- Category 3 (Probably Benign): Initial follow-up suggested.
- Category 4 (Suspicious abnormality): Biopsy should be considered.
- Category 5 (Highly suggestive of malignancy): Appropriate action should be the probability of being cancer.
- Category 6: The biopsy-proven malignancy mammogram.

2.2.4 Challenges in Reading Mammogram

The difficulties faced by radiologists when detecting breast abnormality on mammograms images is based on some factors such as the resolution of the mammogram and the location of the abnormality within the breast tissue. Generally, the smaller the size of the lesion and the more deeply situated within a dense breast tissue the more difficult is its detection by the radiologist. As a result, it might be necessary to take several other images from different views to help the radiologist in making a diagnosis. It is worth mentioning that the younger the patient, the more difficult it is to detect a lesion on a mammogram; this is because younger patients have denser breast tissues, and this will hinder the process of detecting lesions on mammograms. In most of these cases, it would be very important a biopsy taken in order to arrive at the final diagnosis since diagnosis by the mammogram is difficult.

Many statistics show that 30% of breast cancers are undetected (Elshinawy, 2010). Double reading is needed to reduce the rate of missing cancers among radiologists. Double reading means that two radiologists will read the same mammogram. This method has been proven to overcome or reduce the missing rate of cancer but with a major drawback which is the high cost and work load on radiologists.

Because of the aforementioned difficulties, radiologists hold regular meetings to assess their performance in interpreting mammograms. This is conducted via asking a number of radiologists to interpret a set of mammograms, and each radiologist is asked to write his/her report. Reports from all the radiologists are then compared and plotted on a curve to rate the individual performance of each of them. This procedure sheds light on the fact that two different radiologists might comment differently on the same

mammogram especially if the target lesion is small or obscured by dense breast tissue (Elshinawy, 2010) (Bozek, et al., september, 2008).

2.3 CAD Systems

In the following section, we explain the general idea about the Computer Aided-Detection and Diagnosis systems (CAD), followed by CAD architecture of the mammogram system. Finally, we present statistical measurements such as Sensitivity, Specificity, and Accuracy that are used to evaluate the performance of CAD algorithms.

2.3.1 What is a CAD System?

Double reading of mammograms has been applied to reduce the proportion of the misclassified cancers; this is performed by two radiologists read the same mammogram images. However, the workload and cost associated with double reading are relatively high. Instead of double reading, a Computer Aided Detection and Diagnosis system (CAD) is used as a second reader and the final decision is made by the radiologist, which is referred to as the “second pair of eyes of the radiologists”.

The CAD system employs digital image processing procedures to detect a variety of abnormalities in mammogram images including calcifications, masses, and architectural distortion. Expert radiologists can benefit from such procedures that can be of great benefits to early detection of breast cancer which can consequently lower mortality rate among women with breast cancer (Tang, et al., 2009) (Brown, et al., 1996).

The Food and Drug Administration of United states (FDA of U.S) has approved commercially available computer aided detection and diagnosis systems to aid radiologists. Examples of commercial CAD system are iCAD and R2 Technology ImageChecker®. The iCAD is designed to help radiologists and other health care providers detect cancer early. R2 Technology ImageChecker®, is a software application used by radiologists for analyzing two dimensional digital mammography images. This system is designed to search for features that may be associated with breast cancer and mark them (Senthilkumar & Umamaheswari, 2011). Here, an asterisk is used to mark masses while microcalcification clusters are marked with a triangle.

2.3.2 CAD Architecture

Computer-Aided Detection and Diagnosis is designed to provide the radiologist with visual prompts on series of mammograms. When the detection algorithm identifies a suspicious lesion that necessitates further assessment, the system marks those areas on the mammogram to highlight them. This way, the radiologists' interpretations will be complemented. A typical CAD scheme includes the following main stages: Pre-processing, Segmentation, Feature extraction (analysis), and Classification. Figure 2.4 shows the typical processes of CAD system (Masala, 2006) (Costaridou, 2005).

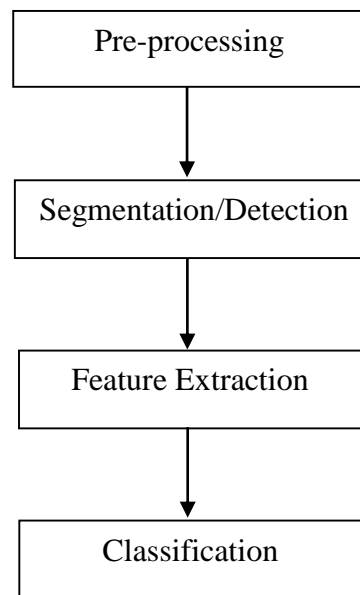


Figure 2.4: CAD architecture

- In the pre-processing step, the subtle features of interest are enhanced and the unwanted characteristics of the image are de-emphasized. The better description of the objects of interest as a result of the enhancement improves the sensitivity of the detection system. In the pre-processing step, the contrast of the region of interest is enhanced, the margins of abnormalities are sharpened, and noise suppression is performed (Costaridou, 2005).
- In the segmentation step, segmentation algorithms have two important principal purposes. The first one is the segmentation methods used to separate the abnormality in the mammogram images from the normal tissue. The second one is the segmentation methods used to split the mammogram image to different components

(background, breast tissue, and pectoral region). The breast tissue is segmented from the image to reduce the amount of image to be analysed by follow-on steps in screening (Sample, 2003).

- Feature extraction is one of the most important parts of the system that effect the CAD performance. Researchers in this field investigate two feature types: (i) those traditionally used by radiologists (gradient-based, intensity-based and geometric features). (ii) high order features that may not be as intuitive to radiologists such as (texture features). An important issue in any CAD system is selecting the best type of features to classify the mammogram images. Taking the single type of feature might not be significant for the classification accuracy but when combined with another type of features might be very significant. The success of classification tasks depends on the use of the most discriminating features. The availability of a large set of features that can be extracted from mammograms necessitates the use of carefully designed feature selection schemes that aim to select an appropriate smaller feature subset of the available set of features, that leads to improved classifier performance (Masala, 2006).
- The classification represents the last step and involves classifying suspicious areas of mammograms into their types such as normal, benign, or malignant tissue. The same classification techniques used in another field of image processing and pattern recognition used for mammograms as well such as Neural networks, Bayesian belief network, Supported vector machine, and K-nearest neighbour. One problem in this manner is how to select the extracted features suitable for various classifiers (Zhang, et al., 2008).

2.3.3 Evaluation of CAD System

The statistical measures of the performance of detection algorithms generally use three indices: sensitivity, specificity and accuracy:

Sensitivity: is the proportion of women positively diagnosed by CAD system as having breast cancer out of those truly diagnosed as having breast cancer by the gold standard test.

$$Sensitivity = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \quad (2.1)$$

Where: True positive (TP) means results obtained by CAD system identifying cancer cases are compatible with the gold standard diagnostic test results, and false negative (FN) means the pathological cases missed by CAD system.

Specificity: is the proportion of women truly diagnosed as not having breast cancer and left alone by CAD system.

$$Specificity = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN) + False\ Positive\ (FP)} \quad (2.2)$$

Where: True negative (TN) means the CAD system results were compatible with the conventional results as not having the disease, and false positive (FP) means cancer cases been identified by CAD system despite the fact that they are not cancer by the gold standard test. Since sensitivity and specificity are proportions, their value will range between 0 and 1; when the value of sensitivity is equal to 1, it means that all cases detected as abnormal are actually abnormal in reality. Similarly, when the value of specificity equals to 1, this means that all cases diagnosed as normal are in fact normal in reality (Shinde, 2003) (Lee, December, 2006) (Yacoub, et al., 2006). We can represent the above terminologies by Table 2.1.

Table 2.1: CAD system evaluation terminology

		Gold Standard	
		Positive (Biopsy demonstrated malignancy)	Negative (Biopsy demonstrated Benign)
CAD	Positive	TP	FP
	Negative	FN	TN

To further explain sensitivity and specificity in the context of CAD system, an example is assumed where we have 1000 women to be tested by CAD system. The result of biopsy shows that 800 of the cases are positive (malignant), and 200 of the cases are negative (benign). The result of CAD system is as shown in Table2.2.

Table2.2: CAD measure example

		Gold Standard	
		Positive (Biopsy demonstrated malignancy)	Negative (Biopsy demonstrated Benign)
CAD	Positive	700	50
	Negative	100	150

Therefore: Sensitivity= $(700/800) = 0.875$, i.e. out of 100 women screened by CAD, 87.5% will be truly positive, and 12.5% will be falsely negative i.e. will receive false reassurance and the disease progress to advance stages before being diagnosed later which adversely affect prognosis.

Specificity= $(150/200) = 0.75$, i.e. out of 100 women screened by CAD, 75% will be truly negative while 25% will be falsely positive. Those false positive cases might receive unnecessary investigation and treatment wasting avoidable resources. Moreover, the diagnosis will put unnecessary stress on the patients and their families (Akobeng, 2007).

Accuracy: is the proportion of true results to the total number of cases (normal and abnormal). It is a testing parameter defined as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (2.3)$$

2.4 Mammogram Image Databases

We need to use a database of mammogram images for a reasonable size population in order to test the performance of any mammogram classification scheme. These tests are conducted according to certain protocols of separating mammograms to testing and training sets for classification purpose that is decided according to one of the mammogram classifiers. There are a number of benchmark mammogram databases that have been recorded for the purpose of experiments; each meets certain criteria regarding structure/kind of the mammograms (Dense, Fatty, or glandular) and (normal, benign, or malignant).

In this thesis, we used the two most commonly used mammogram databases that are publicly available for research purposes; the Mini Mammographic Image Analysis Society (Mini-MIAS) and the Digital Database for Screening Mammography (DDSM). The choice of these databases is based on the fact that, in our application, we face difficulty to collect real cases. On the other hand, we aimed to use mammograms that

are not captured in controlled conditions, and these databases include sufficient noise and low resolution problem, to investigate challenges in such kind of databases for the researchers.

2.4.1 Mini-MIAS Database

The MIAS database is freely available for scientific research purposes (Suckling, et al., 1994). The database consists of 322 mammograms gathered from 161 pairs of Medio-lateral oblique (MLO) views (right and left view). Mammograms originate from a film-screen imaging process in the United Kingdom, National Breast Screening Program (Suckling, et al., 1994). Originally, the MIAS Database images were digitised at 50 micron pixel edge, but were reduced to 200 micron pixel edge and clipped/padded for of research purposes so that every mammogram is of size 1024×1024 pixels resolution with 8-bits per pixel. The new reduced mammogram database called is Mini-MIAS. Depending on the intensity of the mammograms, the database is partitioned into three classes: fatty, glandular and dense. There are 106 fatty mammograms (67 normal and 39 abnormal), 104 glandular mammograms (67 normal and 37 abnormal) and 112 dense mammograms (76 normal and 36 abnormal). Two example mammograms of each class are shown in Figure 2.5.

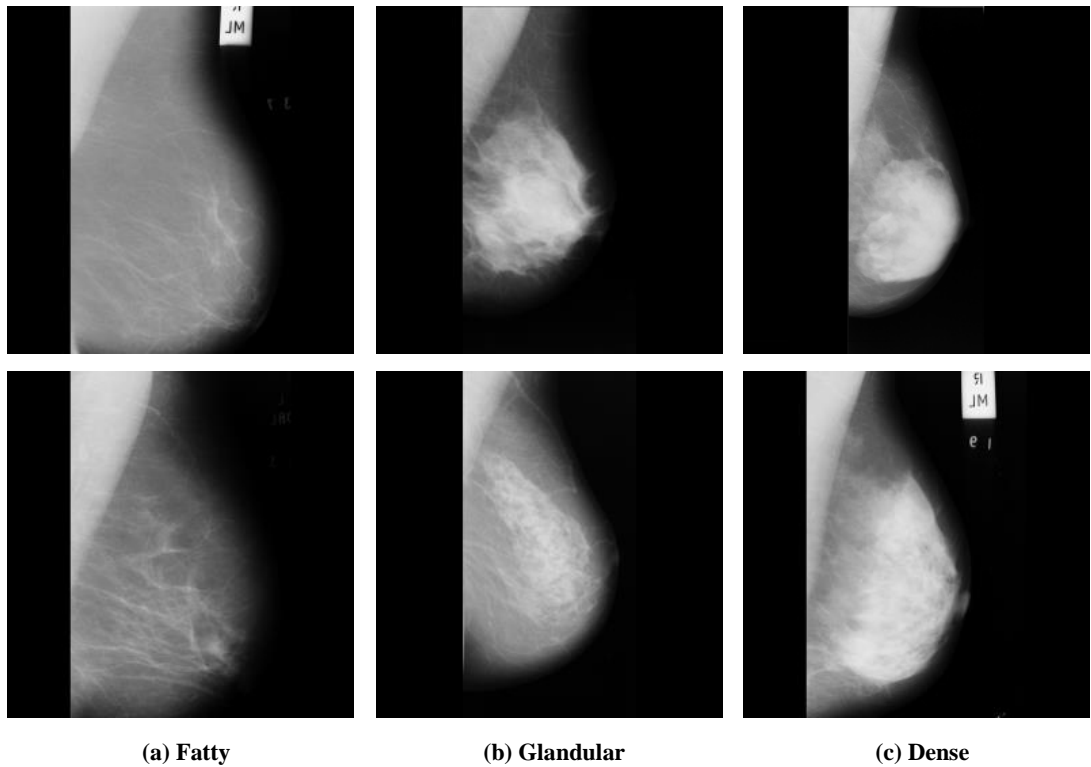


Figure 2.5: Examples of different tissue types available in the Mini-MIAS database

The database includes different types of abnormalities, including calcifications, well-defined Spiculated or ill-defined masses, architectural distortion or asymmetry. The severity of each abnormality is also described, i.e. benignancy or malignancy.

2.4.2 DDSM Database

The commonly used Digital Database for Screening Mammography (DDSM) is a benchmark for testing new proposals dealing with processing and analysis of mammograms for breast cancer detection (Heath, et al., 2001). It contains 2620 cases, each with mammograms captured from four views (Medio-Lateral Oblique and Cranio Caudal) for each case. The mammograms were obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital and Washington University of St. Louis School of Medicine. The database was completed in the fall of 1999. Original images were digitised with three different sampling rates (i.e. 42, 43.5, and 50 micron pixel edges) with different resolutions and two different grey levels (12 or 16 bits per pixel). For experimental purposes, most researches fixed the size of images cropping and converting the grey levels to only 8-bits grey values. Images containing suspicious areas have associated pixel level "ground truth" information about the locations and types of suspicious regions. Some samples of the ROIs are shown in Figure 2.6.

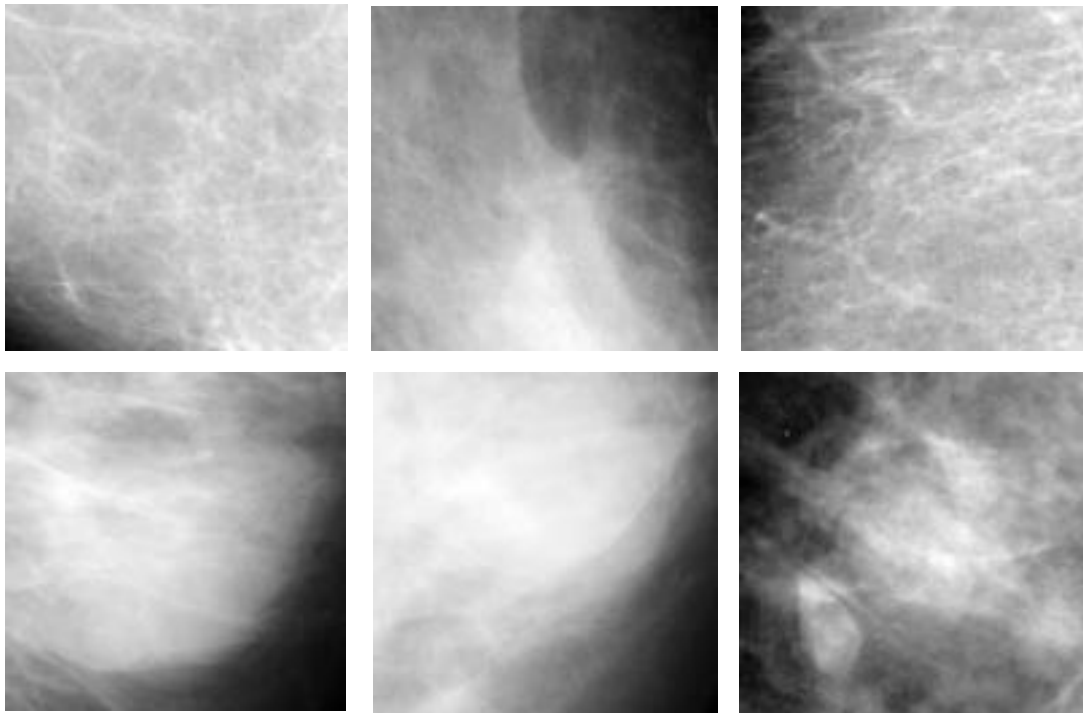


Figure 2.6: Sample mass ROIs first row and normal ROIs second row

Chapter 3

Literature review

Several types of research in the area of computer aided detection and diagnosis systems are available, each of the researches discuss the area depend on their view of the subject. Clearly, it is impossible to include all existing approaches in this chapter. We here focus on publications and works related to our research and provide a brief description of the main points of the approaches. We separated our literature into three main parts. Firstly in Section 3.1 we review CAD breast systems, in general, how CAD systems work and the different problems or limitations they have. Section 3.2 reviews the literature on pre-processing techniques used in automatic mammography systems. This is followed by a review of feature extraction and classification techniques used for mammogram analysis. Finally, the chapter concludes with a brief summary of the overall literature review.

3.1 General CAD System

Computer-aided detection and diagnosis systems play a key role in the early detection of breast cancer and help to reduce the mortality rate and increase the survival rate (Tang, et al., 2009). Subramaniam et al. (Subramaniam, et al., 2006), presented and described a multi-components (referred to as categories) system that they developed for breast cancer diagnosis. The data extraction category discusses methods to improve the process of obtaining breast tissue and record mammogram and MRI (Magnetic Resonance Imaging) in order to aid the medical personal to interpret the images obtained from the breast. They also describe adequate ways of data interpretation that help improving diagnosis. Miscellaneous components are focused on implementing telemedicine in breast diagnosis systems.

Digital image processing techniques play an important role in helping to perform breast biopsies. The role is more for abnormal regions that cannot be felt but can be seen or detected using a conventional mammogram or with using ultrasound. Therefore, before we design a CAD system we need to have knowledge about the differences between

normal breast and abnormal breast tissues. In (Das, et al., 2009) comparison between normal breast tissue and malignant invasive breast tissue was conducted using a series of image processing steps. The comparison also included normal ductal epithelial cells versus ductal / lobular invasive carcinogenic cells. Rangayyan et al. (Rangayyan, et al., 2007) presented an overview of digital image processing and pattern analysis techniques which aimed to address several areas in a CAD system used to detect breast cancer such as: contrast enhancement, detection and analysis of calcifications, detection and analysis of masses and tumours, analysis of bilateral asymmetry, and detection of architectural distortion. In (Bozek, et al., september, 2008) a description of the lesions that could be detected and diagnosed utilizing developed computer-aided detection and diagnosis methods were made. This paper point out further developments is required in CAD algorithms to improve the detection and diagnosis of breast abnormalities using computers.

Masala (Masala, 2006) presents a broad outline of Computer Aided Detection and highlights how CAD systems are correctly used. It also outlines the typical situation where CAD systems can be utilized in the most helpful way. The design of the CAD software for mammography falls within the framework of the MAGIC-5 collaboration. A comparison is made between the newly designed CAD software and the commercial CAD. But (Tang, et al., 2009) is aimed at providing an overview of CAD systems and related techniques developed in recent years, the latest progress that have been achieved in the development of CAD systems and related techniques. Thereafter, the focus will be on the chief CAD techniques developed lately for breast cancer, including detection of calcifications, detection of masses, detection of architectural distortion, detection of bilateral asymmetry, image enhancement, and image retrieval.

In the review paper (Senthilkumar & Umamaheswari, 2011), Computer aided detection and diagnosis are discussed given the mammography-based approach. Here, every technique utilized in each stage of Computer aided detection and diagnosis is explained. This to include the collected information about mammogram databases, CAD systems and also the techniques that have been used to gauge the performance of the systems. In this paper (Zhang, et al., 2008) the author has investigated the most advanced computer aided detection systems for digital mammograms.

The authors in (Sampat, et al., 2008) highlighted the need to improve the previous studies and presented two studies on how to improve the performance of existing

Computer aided detection and diagnosis algorithms. The first study proposed a new evidence-based algorithm for the detection of **spiculated** masses (SM) and architectural distortions (AD), using the statistics of the physical characteristics of these abnormalities to establish the appropriate parameters. The combination of information and features extracted from multiple views was the principle input into the second study while aiming at improving the performance of current diagnosis systems.

It is designed to look at the correspondence of features from multiple views. It attempts to tackle the key issue of whether recruiting data from multiple views can really help to provide further insight to the diagnosis problem. A CBIR system was presented in (Deserno, et al., 2011). In this system, the pattern for image retrieval encompasses the use of breast density along with the existence of a breast lesion. A comprehensive system evaluation was continued based on a considerably enlarged database to include 3,375 mammograms of totally 12 classes. This work contributes to (i) CBIR-CAD of mammography, (ii) producing a system that can assist radiologists in their diagnosis, or (iii) useful as a pre-processing step for CAD systems for breast lesions classification.

From above overview, we had seen that until now there is not any general CAD system that works perfectly without any errors. Therefore, we provided a brief review here of the published reviews of different CAD techniques for breast cancer detection and diagnosis specifically.

In summary, a variety of different algorithms and techniques have been used to improve the performance of CAD systems, and no specific techniques have proved to work better than other.

3.2 Pre-processing (Segmentation)

As a pre-processing step for any automatic mammography image analysis, the mammogram is segmented in order to reduce the search zone for abnormalities without undue influence from a background of the mammogram image. As we mentioned in chapter two, the advantages of using the segmentation processes include reduction of noise and edge-shadowing effect, accurate detection of ROI for pectoral muscle, and suppress the pectoral muscle successfully. This way, the processed mammogram can be used for the automated detection of human breast abnormalities such as calcification, masses (circumscribed, spiculated, ill-defined), asymmetry analysis... etc.

The pectoral muscle is a high-intensity that must be removed. Other image pre-processing procedures include the removal of different types of noise, high intensity rectangular label, together with low intensity label and tape artefacts (Tzikopoulos, et al., 2011). Pectoral muscle appears at approximately the same density as the dense tissues of interest within the breast border in the mammogram image. Therefore, segmentation of pectoral muscle is important to limit the search for the relevant breast abnormalities. Evaluating the success of segmentation and classification techniques, are dependent on using available mammography image databases, especially Mini-MIAS (Suckling, et al., 1994). Traditional segmentation method for breast border contour detection was done manually or semiautomatic (Chandrasekhar & Attikiouzel, 1996) using devices such as a computer mouse. These methods are tedious and undesirable in today's practices.

Earlier methods for automatic identification of the pectoral muscle were based on the use Hough transform proposed by Karssemeijer (Karssemeijer, 1998). The Sobel operator was used to calculate the gradient image within a triangular ROI extracted. Then the Hough transform was performed on the gradient image to detect the pectoral boundary. The Hough space can be viewed as an accumulator array of parameter values representing all possible straight lines, with peaks (i.e. local maxima in the region) corresponding to straight lines in the image. A threshold is used to eliminate all elements of the accumulator array that are not a local maximum in this window. From the remaining candidate points in the area of the pectoral muscle, segmented by the corresponding line in the image, only one peak was selected to be inversely transformed into the image space and all pixels above the line in the ROI were marked as the pectoral muscle. When tested for 615 digitized mammograms, exact agreement with the manual process was obtained in 65% of the cases.

Kwok (Kwok, et al., 2004) proposed a new method consisting of two major components. First, the pectoral muscle boundary was approximated with a straight line, which was then refined to a curve using an iterative cliff detection. The straight line was estimated within the ROI based on iterative threshold selection and straight line fitting incorporating a gradient test. The resulting line was validated to adjust the ROI and perform the second estimation. The new straight line was then used to initialise the cliff detection, where the intensity profiles along a set of search paths perpendicular to the straight line were analysed, and the cliff locations were detected by fitting the

smoothed profile to a sigmoid function. This procedure was performed iteratively to obtain a more accurate curve. Finally, a closed region was generated to segment the pectoral muscle. This method was adaptive to large variations in edge strength and texture of the pectoral boundary. In addition, it remained effective when the pectoral boundary was partially obscured by overlaid glandular tissue or artifacts. Two radiologists assess the result of the segmentation method, radiologist 1 rated 75.5% images as acceptable for the straight line segmentation and 87.0% images for curve segmentation, while radiologist 2 rated 52.8% images for straight line segmentation and 67.1% for curve segmentation. In the same year, Ferrari (Ferrari, et al., 2004) presented two methods, the first method depends on Hough transform to detect the pectoral muscle and the second method based on the Gabor wavelet filter bank. For enhancing the pectoral muscle edge with the region of interest containing the pectoral muscle, the scheme starts by convolving a group of Gabor filters which are specially designed for this purpose. The technique tested on the set of 84 MLO mammograms from the Mini-MIAS database, the accuracy of the method evaluated by using Hausdorff Distance Measure HDM for the first method HDM is equal to 7.08 ± 5.26 and for the second method, HDM is equal to 3.84 ± 1.73 . We must note that the second method proposed to overcome the limitation of the hypothesis of a straight-line representation of the pectoral muscle for the first method (Ferrari, et al., 2000), which was a modification of the method proposed by Karssemeijer (Karssemeijer, 1998).

In contrast to the edge based methods, Raba (Raba, et al., 2005) present a region growing technique for segmenting a digital mammogram into breast region and background with pectoral muscle suppression. An intensity threshold estimation was used to locate the pectoral muscles in the mammogram. Firstly, the breast orientation was identified by detecting the curvature on each side of the breast. Secondly, a seed point was placed at the first pixel of the non-curved side. Finally, the region growing started from this seed point to fill the pectoral muscle based on an intensity criterion. A constraint on the size of the grown region was applied to avoid an overgrown region. The technique tested on the set of 320 mammogram images from the Mini-MIAS database; the results are obtained from a visual inspection by experienced radiologist they obtained 86% of good extractions for the muscle subtraction. We should note that for the dense tissues which appear near the pectoral muscle, the region-growing algorithm may produce segmentation leakage in which dense tissues are included in pectoral muscle region, for this reason, we should notice that some of the results are a

little bit over or under segmented. Also in (Mirzaalian, et al., 2007) a new method was proposed for detecting the pectoral muscle in MLO mammograms using nonlinear diffusion algorithm. The algorithm was tested on only 90 mammogram images selected from Mini-MIAS database. The result shows that this method outperforms the method in (Kwok, et al., 2004). A hybrid method for the pectoral muscle detection using bit depth reduction and wavelet decomposition was proposed by Mustra (Mustra, et al., 2009). Only 40 mammograms are used to test the method. The results of pectoral muscle segmentation are divided into three groups Good 47.5%, acceptable 37.5% and unacceptable 15%. The work in (Nagi, et al., 2010) uses morphological operations as a pre-processing and seeded region growing (SRG) method used to detect and delete noises, suppress artifacts, separate background region from the breast region, and remove the pectoral muscle.

Recently, detection of pectoral muscle boundary based on a graph-cut-based image segmentation technique and Bezier curve are presented by Camilus (Camilus, et al., 2010), the method evaluated by using 84 images from mammographic image analysis society database. The mean false positive and false negative rates are respectively as 0.64% and 5.58%. Authors in (Chen & Zwiggelaar, 2010) proposed a method based on histogram thresholding, edge detection in scale space, contour growing, polynomial fitting and region growing. The method tested 240 mammograms from EPIC database; the result were visually rated as four categories: accurate, nearly accurate, acceptable and unacceptable. For the breast background segmentation, 66.5% are accurate, 25% are nearly accurate, 6.9% are acceptable, and 1.6% are unacceptable. For the pectoral muscle removal, we have obtained 62.5% accurate, 25.4% nearly accurate, 5.6% acceptable, and 6.5% unacceptable segmentations. Also in (Vaidehi & Subashini, 2013) authors proposed a method to remove the pectoral muscle depend on drawing a straight line then using a threshold to detect the pectoral muscles, the method evaluated by using 120 images from the Mini-MIAS database, the accuracy was 85% were compared with manually drawn by the radiologist. **The work by** (Vikhe & Thool, 2016) proposed an intensity based approach for pectoral muscle boundary detection in mammograms, the simple filter mask of 3×2 has been applied with few coefficients to enhance the pectoral region. The significance of the designed filter is, it considers the direction along with transition gray level intensity changes across the pectoral region compares to conventional enhancement filter. After pectoral enhancement, row and column wise search approach has been applied to the candidates to obtain the pectoral

boundary points, based on threshold technique (black dots). The boundary points are finally connected to detect the pectoral muscle. The method was tested on 320 digitized mammograms from Mini-MIAS database. The result was visually rated as three categories: accurate 87.19%, acceptable 9.37%, and inaccurate 3.44%.

All the aforementioned papers focused on obtaining the image of the breast tissue without the pectoral muscle using different techniques to remove the pectoral muscle. Not only they have used various techniques to remove the pectorals, but they also have used different databases and chosen a different number of images as their random sample. It can be noted that there is a common diagnostic problem that is encountered in these studies, that is, the problem of a dense breast tissue. This is because a dense breast tissue has a radiological intensity that is similar to the intensity of the muscle making the process of distinguishing between the two troublesome.

The rather modest accuracy rate for the above methods, coupled with the complexity of their implementation, was the motivation to propose and design new methods to remove the pectoral muscle, extract the breast border, and remove labels and artefacts. The new methodology we adopted in (Majeed, et al., 2013) has the advantage of recruiting all the images available on Mini-MIAS database, thus accounting for the higher accuracy rate, without excluding any images for the purposes of breast border extraction and pectoral muscle removal for more details see Chapter 4 of this thesis. Table 3.1 shows the summary of the pre-processing (segmentation) techniques available in the field.

Table 3.1: Pre-processing (segmentation) techniques summary

Author & Year	Approach	Database	Performance
(Karssemeijer, 1998)	Sobel operator and Hough transform	615	571 a pectoral muscle was segmented
(Ferrari, et al., 2000)	The Hough transform and edge detection	84 Mini-MIAS	FP= 1.98±6.09%, FN=25.19±19.14%
(Ferrari, et al., 2004)	The Hough transform and Gabor filters	84 Mini-MIAS	HDM= 7.08±5.26 and 3.84±1.73
(Kwok, et al., 2004)	Edge detection using cliff detection	322 MIAS	83.9% adequate
(Raba, et al., 2005)	Region growing	320 Mini-MIAS	86% good
(Camilus, et al., 2010)	Curve fitting and graph theory	84 MIAS	FP=0.64%, FN=5.58%
(Chen & Zwigelaar, 2010)	Histogram thresholding, edge detection, contour growing, polynomial fitting and region growing	240 EPIC	66.5% accurate, 25% nearly accurate, 6.9% acceptable, and 1.6% unacceptable for background segmentation, and 62.5%, 25.4%, 5.6%, and 6.5% for segmentation respectively
(Vaidehi & Subashini, 2013)	Straight line and threshold	120 Mini-MIAS	85%
(Vikhe & Thool, 2016)	intensity based approach and threshold	320 Mini-MIAS	87.19% accurate, 9.37% acceptable and 3.44% inaccurate

3.3 Feature Extraction and Classification

Feature extraction is an important part of the final classification step. Classification is a process of discriminating between two or more different classes by labeling each similar set of data in order to differentiate it from other classes. It is one major task in machine learning.

There is a relation between feature extraction and the classification process, therefore, the higher accuracy of the classifier can be achieved by selecting the optimum feature set. During our study, we found that there are several types of research in the area, and several types of classifier used to classify the mammogram also the type of the database that used is different from study to another. So we will explain and discuss the recent state of the art of the feature extraction and classification which is related to our work. There is no specific single set of features known to produce optimum results for

classifying mammograms. The reason behind this is the irregularity and randomness of the abnormal mammogram patterns. The main challenge in feature extraction is the ability to find the set of features that best suits each type of tissue for mammograms.

Textural features are widely used to extract features from the mammogram and then classify the mammogram. Different approaches based on co-occurrences matrix as a feature descriptors extraction were developed. Bovis and Singh (Bovis & Singh, 2000) used five statistical features obtained from co-occurrence matrices extracted from four different spatial orientations, horizontal, left diagonal, vertical and right diagonal corresponding to (0, 45, 90 and 135) degree and four pixel distance ($d=1, 3, 6, 9$). The method was applied on 144 mammograms from MIAS database and the classification accuracy obtained using Artificial Neural Network classifier.

In 2003, Youssry (Youssry, et al., 2003) used a neuro-fuzzy model to detect the circumscribed masses in mammogram images. The texture feature was estimated using co-occurrence matrices, and feature selection on the correlation coefficient between features was performed. The selected features were used to train a neuro-fuzzy model. The method was applied to small set of mammograms (twenty two mammograms) from the MIAS database (i.e. eleven fatty, eight glandular, and three dense). On the other hand, Marti (Marti, et al., 2003) presented a supervised method to segment the mass in the mammogram image. Texture features were used to present a homogeneous behaviour inside the selected region. Karahaliou (Karahaliou, et al., 2008) presented a study to explore texture properties of the tissue surrounding microcalcification (MC) clusters on mammograms for breast cancer diagnosis. The analysed mammograms include 100 MC clusters (46 benign and 54 malignant). Using redundant discrete wavelet transform, tissue surrounding MCs is defined on wavelet decomposed images. Gray-level texture and wavelet coefficient texture features at three decomposition levels are extracted from Surrounding Tissue Regions of Interest (ST-ROIs). The ability of the method in differentiating malignant from the benign tissue is investigated using a probabilistic neural network. Also, Elshinawy (Elshinawy, 2010) used Gray level co-occurrence matrix as texture features to developed a pre-normal CAD algorithm that will identify the characterization of normal mammograms in each tissue type (dense and fatty). She concludes that GLCM is a good feature for fatty-tissue mammograms. For the purpose of differentiating the mammogram to malignant and benign Mohanty (Mohanty, et al., 2011) extract nineteen features from grey level run-

length matrix and grey level co-occurrence matrix to distinguish between malignant masses from the benign mass. The work of Fatima Eddaoudi (Eddaoudi, et al., 2011) focused on the detection of masses in the mammogram image by using SVM classification and textures analysis. The identification of tumours is generally made in three stages: (i) pectoral muscle segmentation, (ii) hard density zone detection and (iii) texture analysis of regions of interest. The region of interesting segmented is classified using Haralick features calculated from the co-occurrence matrix.

The discrete wavelet transform (DWT) is a multi-resolution image analysis tool used in many image processing and computer vision applications, including automatic mammogram analysis. For example, Dong and Wang (Dong & Wang, 2009) used Gabor filters of different frequencies and orientations to extract textural patterns. The scheme begins by producing N filtered outputs from N Gabor filters with different orientations and frequencies. Then for each individually filtered output, the mean and the standard deviation of the coefficient magnitude are calculated and used as image features. Thirdly, a statistical feature selection t-test is applied to each feature and tests the p-value. If the p-value is less than 0.05, then the feature is kept, otherwise, it is discarded. Finally, K nearest neighbours used to classify the mammograms into normal or abnormal. MIAS database used for the evaluation process.

In the same year, Eltoukhy (Eltoukhy, et al., 2009) proposed a method depend on extracting an amount of the biggest coefficients in each level of a multilevel decomposition. Three different wavelets are tested Daubechies-8 (db8), (sym8), and (bior3.7) wavelets, a set of coefficient from each level of decomposition is extracted to used it to classify the mammogram to normal or cancer and also classify the abnormal class to benign or malignant. In each level of decomposition, the biggest 100 coefficients are used to represent the feature vector. Also, 100 coefficients are extracted from each level of decompositions of the three functions, then Euclidean distance method used to the classification process. Diagnosing the breast cancer from mammogram images was developed by Naveed (Naveed, et al., 2011). The author presented the diagnosis system in different phases: (i) pre-processing on mammogram image is done by using fuzzy filter to remove the noise, background removal, and pectoral muscle separation (ii) discrete wavelet transform is used for feature extraction (iii) different classifiers are used to classify the mammogram to malignant and benign (iv) the malignant images are again classified using one against all technique to find

abnormalities present in the mammograms. The Mammographic Institute Society Analysis (MIAS) database was used for the experiment, and another real-time dataset is taken from Shaikat Khanum Memorial Cancer Hospital and Research Centre (SKMCH and RC) which only contains records of 80 cases (37 benign and 43 malignant). On the other hand, a classification approach of mammographic microcalcifications was presented based on texture by Dhawan (Dhawan, et al., 1997). They extract the global texture based feature from Gray level co-occurrence matrix, and the local texture feature is extracted from wavelet packets gained by decomposing the region at the first level of decomposition. The radial-basis-function neural network classifier was found to be satisfactory especially for cases where the diagnosis was difficult. The results revealed that the neural network is a robust classifier of a combination of local and global texture features into benign and malignant groups.

Kramer and Aghdasi (Kramer & Aghdasi, 1998) classify the microcalcifications in mammogram images by using a multiscale statistical texture classification system. The system utilizes the multiresolution analytical tools provided by the wavelet transform in conjunction with the statistical textural information extracted from the co-occurrence matrix. These co-occurrence texture signatures are used alongside other wavelet texture signatures to classify the microcalcifications as benign or malignant. In the study, a comparison was made between the two feature sets, and it was found that the two features perform equally well when the two types of texture signatures are extracted from a decomposition based on the biorthogonal wavelet. The co-occurrence texture signatures outweighed the wavelet texture signatures when using the Daubechies wavelets for the decomposition. A limited number of cases in the database were used for experiments.

Fathima et al. (Fathima, et al., 2013) proposed a method based on extracting first order textural features, GLCM features, DWT features, run length features and higher order gradient features from the ROI. SVM was used as a classifier to separate benign and malignant cases. Local binary patterns (LBPs) have been successfully used as an image texture analysis tool (Ojala, et al., 1996) (Ojala, et al., 2002). The LBP operator was proposed by Ojala (Ojala, et al., 1996) and soon gained popularity as a powerful texture descriptor for application in texture analysis, motion detection, image retrieval, remote sensing, biomedical image analysis and face recognition. Among all above mentioned applications, LBP scheme has shown its potential in face recognition (Guo, et al., 2010)

under different illumination and expression conditions and it is one of the most popular local feature-based methods. The application of LBP in face recognition was first proposed by Ahonen (Ahonen, et al., 2004).

More recently, multiresolution LBPs have been shown to provide highly discriminative texture information in a number of image processing and pattern recognition applications. A recent survey of LBPs and its variants is presented by Huang (Huang, et al., 2011). Advantages of LBPs are said to be its invariance against monotonic gray level changes and its computational simplicity, which makes it possible to use LBPs to analyse images in challenging real-time applications. For example, Oliver et al. (Oliver, et al., 2007) used LBPs to separate truly recognised masses and the ones which actually are normal parenchyma. The method starts by dividing the ROI images into small regions and computes local texture feature using LBPs. Extracted local descriptors are combined in a spatially enhanced histogram to produce a final feature vector descriptor. Finally, the feature descriptors are used to classify the ROI mammograms into true masses and normal parenchyma.

For the purpose of detecting the normal mammogram based on the density (fatty and dense) of the breast, Elshinawy et al. (Elshinawy, et al., 2010) used three types of LBP-based features (simple LBP, multi-resolution gray-scale with rotation invariant texture classification in LBP, and LBP histogram on Fourier features). SVM classifier was used for the classification purpose. Also, Berbar et al. (Berbar, et al., 2012) proposed a method based on two different types of features which are statistical features and LBP features for classification of abnormality in digital mammograms. Yanfeng et al. (Yanfeng, et al., 2015) used k-nearest neighbour (KNN) classifier to classify mass region to benign or malignant cases. The authors proposed a method by combining texton analysis with multiple sub sampling strategies to classify 144 mass regions from Digital Database for Screening Mammography (DDSM) database, the accuracy rate to benign or malignant. The classification accuracy rate is 85.96%.

The main challenge in feature extraction is the ability to find the set of features that best suits each type of tissue for mammograms and all types of abnormality. A primary difference in our work with the above mentioned techniques is that we focus on separating the mammogram to normal and abnormal in different tissue type (Fatty, Glandular, and Dense) and different types of abnormality (Mass, Microcalcification, ..., etc). Details of our proposed feature extraction methods will be described in Chapter

5. Based on the performance of our proposed features and our review of existing work here, we will show that there is no specific feature representation that is suitable for all kind of mammogram cases. Therefore, we propose the use of fusion as a means to improve mammogram classification accuracy. The fusion will be performed in two ways: (i) feature fusion (i.e. fusing different feature sets extracted from the same mammogram image into one larger feature vector), then the new feature set will be presented to a single classifier to obtain final classification result; (ii) decision level fusion (i.e. different sets of features extracted from the same mammogram are presented to a classifier, one set at a time. Then the results of each classification result will be fused to obtain final classification result). The proposed use of fusion is presented in Chapter 7. Our automatic models aim to classify the images to normal and abnormal, and by doing so, aims to improve the radiologist's performance by screen-out the normal mammograms as a first-look and therefore leave the radiologists with the suspicious cases for more analysis. The summary of available mammogram feature extraction and classifications are shown in Table 3.2.

Table 3.2: Feature extraction and classification summary

Author & Year	Approach	Database	Classifier
(Bovis & Singh, 2000)	Statistical feature, co-occurrence matrix	144-MIAS	ANN
(Yousry, et al., 2003)	Texture feature, co-occurrence matrix	22-MIAS	Neuro-fuzzy model
(Karahaliou, et al., 2008)	Gray-level texture and wavelet texture feature	100 mammogram	Neural network
(Elshinawy, 2010)	GLCM matrix, ,LBP	DDSM	SVM
(Mohanty, et al., 2011)	Gray level run-length matrix and GLCM	88 mammogram	Association rules mining
(Eddaoudi, et al., 2011)	Texture feature from GLCM matrix	MIAS	SVM
(Dong & Wang, 2009)	Gabor filter, mean and standard deviation of the coefficient magnitude	MIAS	KNN
(Eltoukhy, et al., 2009)	Wavelet transform (db8), (sym8) and (bior3.7)	MIAS	KNN

Author & Year	Approach	Database	Classifier
(Naveed, et al., 2011)	Discrete wavelet transform	MIAS, (SKMOLT and RC)	Ensemble classifier
(Dhawan, et al., 1997)	GLCM, wavelet	191 mammogram	NN
(Kramer & Aghdasi, 1998)	Multiscale statistical texture, wavelet transform, GLCM	104- Nijmegen	K-NN
(Fathima, et al., 2013)	GLCM feature, DWT feature, run length feature, higher order gradient feature	MIAS	SVM
(Oliver, et al., 2007)	LBP	DDSM	SVM
(Elshinawy, et al., 2010)	Simple LBP, multi-resolution gray-scale with rotation in LBP, and LBP on Fourier feature	DDSM	SVM
(Berbar, et al., 2012)	Statistical feature and LBP feature	512- DDSM	KNN, SVM
(Yanfeng, et al., 2015)	Texton analysis with multiple sub sampling	144-DDSM	KNN

3.4 Summary

In this chapter, we reviewed and presented a number of existing mammogram image analysis approaches. We started by reviewing general CAD systems, how CAD systems are work and what their problems are. Then we reviewed the literature on pre-processing and mammogram segmentation.

The aim of pre-processing and segmentation is to obtain the breast tissue of mammogram image without the pectoral muscle, labels, and artifacts. The next chapter is dedicated to our work on pre-processing and segmentation. We aim to design an accurate method of breast border extraction, artifacts removal and removal of annotations typically found in the background of the mammogram images. Our method will be tested on all images available in the Mini-MIAS database. This work has been published here (Majeed, et al., 2013).

In addition to pre-processing and segmentation, in this chapter, we reviewed existing literature on feature extraction and classification. As a conclusion of the review, we

can say that; till now there is no specific single set of features known to produce optimum results for classifying mammograms. The reason behind this is the irregularity and randomness of the abnormal mammogram patterns. The main challenge and open problem for the researchers are to find the set of features that best suits each type of tissue for mammograms. Our aim is to classify the images to be either normal or abnormal. Therefore, we will investigate the performance of several texture features and propose the multiscale texture features for mammogram classification. This work will be presented in Chapter 5. Moreover, to realise an efficient classification system, we apply the idea of dimension reduction on the original mammogram feature vectors (more details can be found at the end of Chapter 5).

Based on our findings in Chapter 5, we propose the use of fusion to enhance mammogram classification accuracy further. This is analogous to more than one radiologist inspecting the same mammogram to arrive at a decision. The fusion will be done in two ways; feature level fusion and decision level fusion. Chapter 6 is devoted to the use of fusion in mammogram classification.

Finally, we consider the use of image super-resolution to enhance mammogram images prior to extracting texture features. This work will be presented in Chapter 7.

Chapter 4

Breast Border Extraction and Pectoral Muscle Removal in MLO Mammogram

The previous chapter presented various exiting techniques for mammographic pre-processing and segmentation that aim to obtain the breast tissue of a mammogram without the pectoral muscle, labels, and artifacts. In this chapter, we initiate our investigations and development of accurate schemes to breast border extraction, label removal, and pectoral muscles removal. We will first describe, in Section 4.1, the need for breast region segmentation. In Section 4.2 we propose a new scheme to extract the border of the breast and remove noise and artifacts. In Section 4.3 we describe in details the proposed pectoral muscle removal scheme. The experimental setup including the database (Mini-MIAS database) used, experiment strategy (Leave-one-out strategy), the feature vector, and classification types are described in Section 4.4. Section 4.5 includes the discussion of the results obtained using mammogram classifiers as well as the expert radiologists. The chapter will end with concluding remarks in Section 4.6.

4.1 Breast Region Segmentation

In order to automatically analyse a digitised mammogram by a computer, the mammogram should be segmented into its representative anatomical regions. The segmentation is very important for a number of reasons but mainly because it reduces the search zone for abnormalities and makes it limited to the relevant region of the breast without undue influence from the background of the mammogram, which typically includes annotations and artifacts that could have an adverse effect on the analysis (Bandyopadhyay, 2010) (Raba, et al., 2005).

In the Medio-Lateral Oblique view, the pectoral muscle always appears as a high intensity, triangular region across the upper posterior margin of the mammogram. Types of noise observed in mammograms include high intensity rectangular label, low

intensity label and tape artifacts (Tzikopoulos, et al., 2011). A typical mammographic image shown in Figure 4.1 illustrate the different types of high level of noise and the artifacts that have to be removed.

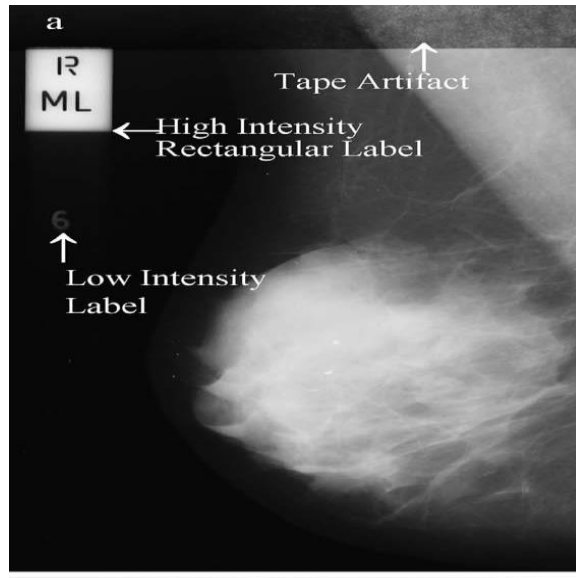


Figure 4.1: Types of noises example observed at a mammogram mdb002 in Mini-MIAS database (Suckling, et al., 1994)

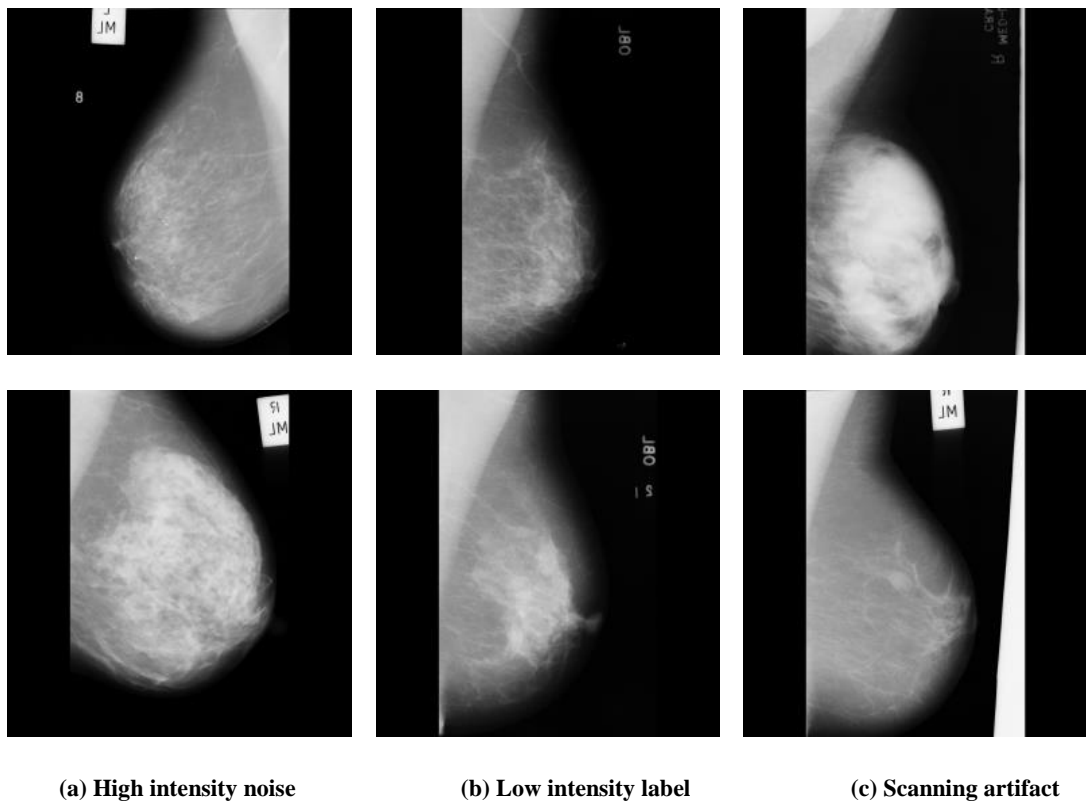


Figure 4.2: Example images of the Mini-MIAS database with different types of noises and background artifacts

Figure 4.2 shows extra example images from Mini-MIAS Database to justify the need for the pre-processing step before extracting features for classification. We can observe three of the main types of noise that need to be removed from the mammograms: (a) high intensity noise regions including the bright rectangular labels, (b) low intensity labels and (c) high intensity scanning artifacts. These examples highlight the pectoral muscle as well.

4.2 Breast Border Extraction

There are two major problems in the accurate identification of the breast boundary of the mammogram image. (i) The region near the breast boundary has decreasing contrast, caused by a lack of uniform compression of breast tissue during the acquisition process. For digitized mammographic images, the digitisation process may even further decrease the visibility of the breast boundary due to additional noise. The low visibility of the breast skin line, makes the identification of the breast boundary so difficult. (ii) The other problem is the non-uniform background that may contain high-intensity regions, such as annotations, labels, frames, and unexposed regions, which interferes with the segmentation of the breast region (Wirth & Stapinski, 2003).

To deal with the aforementioned problems, and to develop more accurate border extraction techniques compared to the techniques discussed in the previous chapter, we propose a new method to extract the breast border and remove noise and artifacts. The overall process of the proposed scheme summarized as follows: in the first step we binarise the input image using adaptive thresholding scheme (i.e. the threshold value is selected based on the intensity of the input image) where the $\text{Threshold} = \text{Mean of the image intensity} - (\text{Mean of the image intensity}/3)$. The output of this step is the binarised mammogram image that consists of several objects including artifacts and the breast region. Binarised objects are then disconnected by applying morphological erosion with structure element = $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, to obtain a number of independent binarised objects. Then we retain the largest object (which is the breast region) inside the image and delete all other small objects. Subsequently, we isolate the spiky boundary (near the edges of the image) and smooth the remaining region by applying morphological dilation process with structure element of 15 by 15 matrix which all elements are ones. By now, a binary mask is obtained consisting of the breast image and the background. Finally, the binary mask is used as an index indicator to extract

the original pixel value of the image. The output of this process will consist of the breast region including the pectoral muscle. The overall process is shown in Figure 4.3.

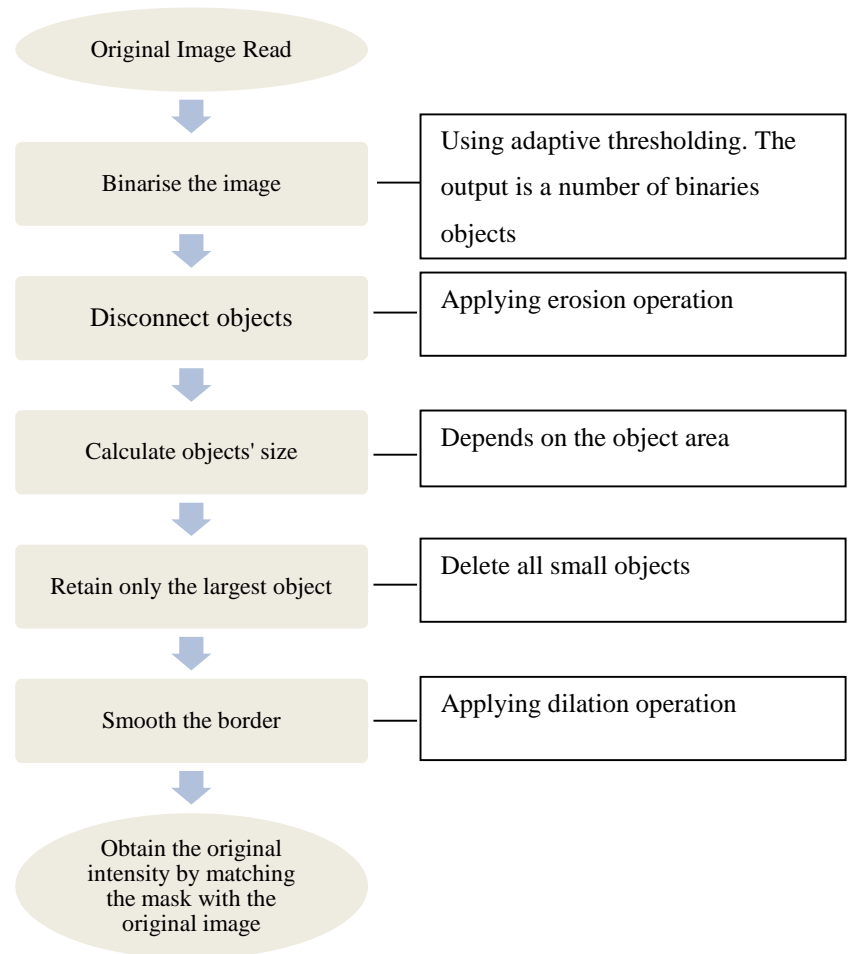


Figure 4.3: Breast border extraction process

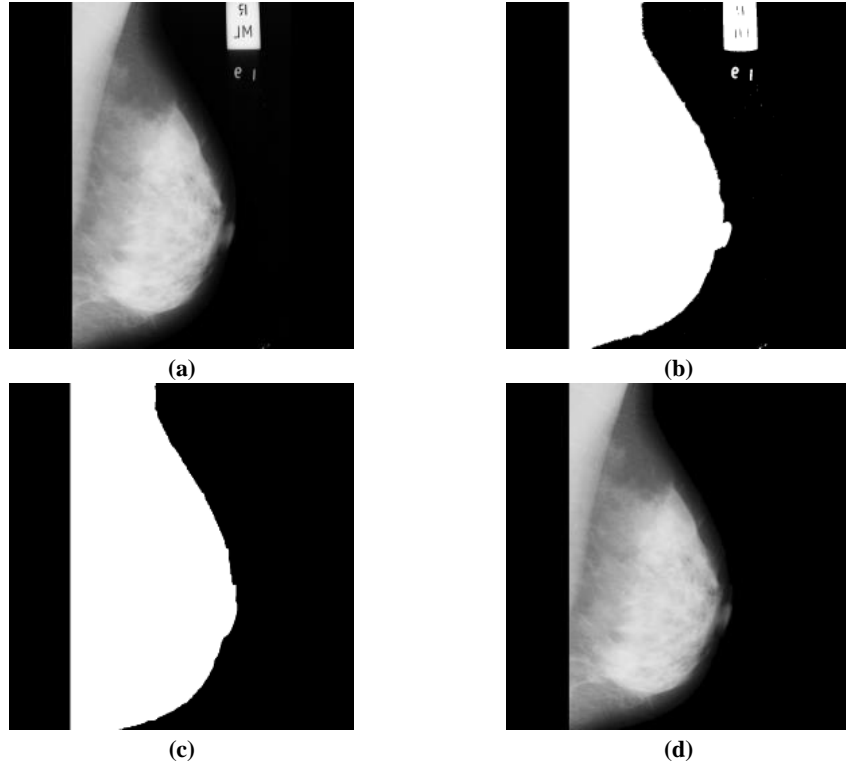


Figure 4.4: (a) Original image, (b) Binarised image, (c) Removed artifacts, (d) Extracted breast region

Figure 4.4 shows the output images obtained from each step of the proposed breast border extraction scheme explained above.

4.3 Pectoral Muscle Removal

The pectoral muscle represents a predominant density region in most Medio-lateral oblique (MLO) views of mammograms. Automatic identification of the pectoral muscle is useful in many aspects of mammographic image analysis. The exclusion of the pectoral muscle is necessary to avoid bias in the detection of breast cancer. As we described in Chapter 3 Section 3.2, many methods for breast region segmentation have been developed based on various image processing techniques.

However, it is still a difficult task to achieve accurate segmentation of the breast region in particular for digitised mammograms, due to the presence of severe noise and the non-uniform background. We propose a new method for pectoral muscle removal to achieve an improvement in overall classification performance.

The pectoral muscle that remained in the image after the previous segmentation steps (Section 4.2) appears at the upper-left corner (all images are mirrored to the left) of the image with high intensity when compared with the other breast regions. This image

will be used as an input to the proposed algorithm to remove the pectoral muscles. The overall process is shown in Figure 4.5 and the output of each step shown in Figure 4.6.

The rather modest accuracy rate for the above methods, coupled with the complexity of their implementation, was the motivation to propose and design new simple methods to remove the pectoral muscle, extract the breast border, and remove labels and artefacts. Our proposed method is a snake-like method that traces the visible thin line separating the breast tissue from the pectoral muscle. This thin line detection is based on an adaptive threshold chosen in terms of the differences between pixel intensities on both sides of the line. For removing labels, artefact, and breast border extraction this is done in one go using the morphological operations and size based objects detection explained in section 4.2 above. The detail of each individual steps of the proposed algorithm can be explained in steps as below:

Step 1: Crop the black left part from the image. This step is done by checking the intensity value of the obtained image from the previous scheme. If the intensity value is zero, then we will ignore the checked column and look to the next column. This process will continue until the algorithm reaches a pixel with an intensity value greater than 0.

Step 2: Starting from the top row, move downwards by performing the following:

- a. Select row (x) position and scan columns(y). When the pixel value is less than a given suitable (chosen) intensity value, then save the column (y) index. Chosen threshold = maximum intensity value - 1/3 of the maximum intensity value).
- b. If column index is less than a given threshold (i.e. the column is very near to the left edge of the image; experimentally we choose 5 as a threshold) then go to step 3, otherwise, increase row position by a specific amount and repeat step 2.

Step 3: After Step 2 is completed we will obtain a number of (x,y) coordinates representing points on the muscles boundary.

Step 4: Draw a line between neighbouring points that were found from Step 2 to get a longer line that will identify the muscle boundary.

Step 5: Assign 0 to all pixel values on the left of the muscle boundary.

The output of this algorithm is the breast region after the pectoral muscle is removed.

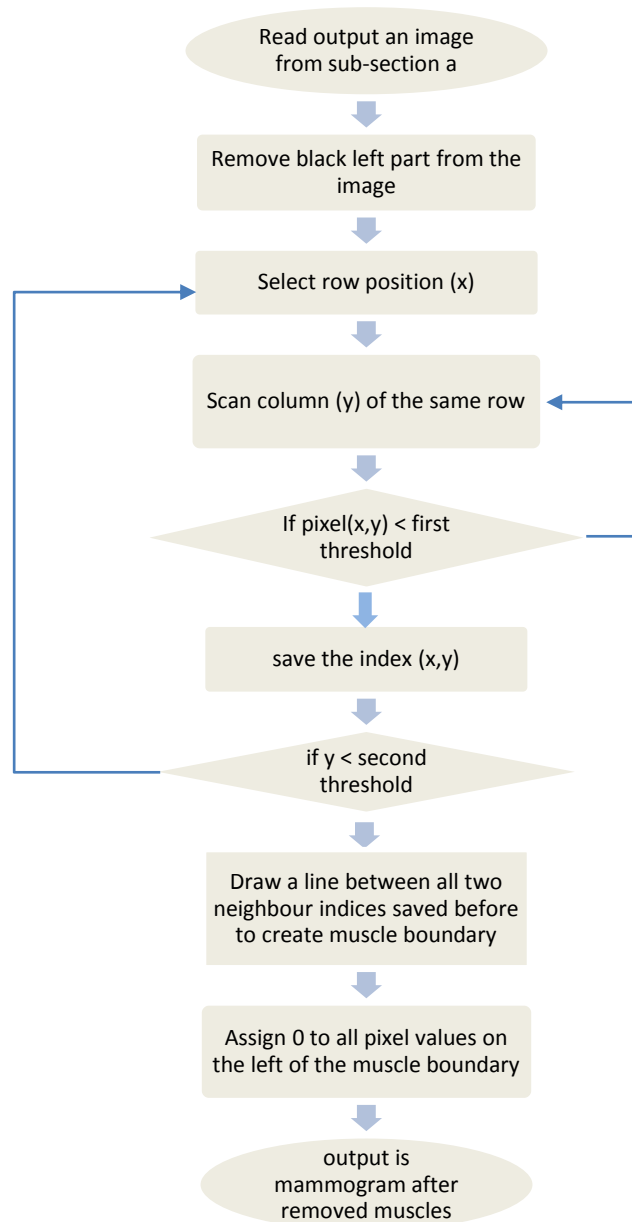


Figure 4.5: Pectoral muscle removal process

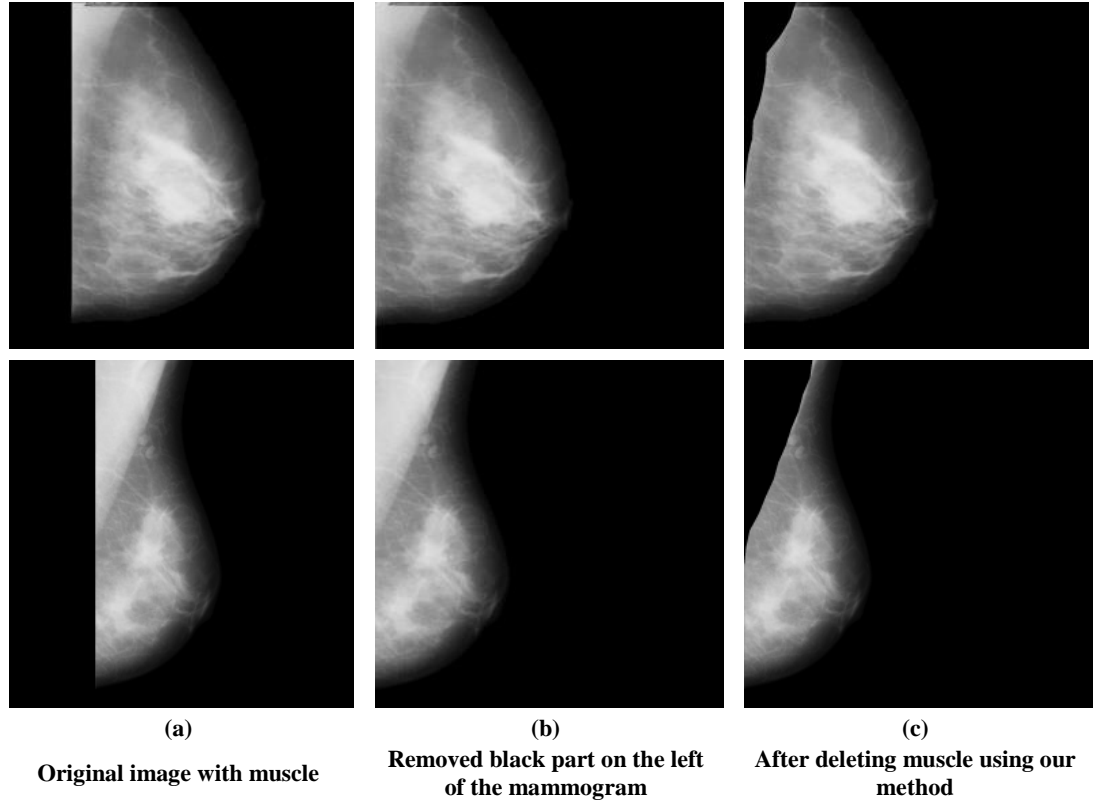


Figure 4.6: Output of the Pectoral muscle removal algorithm

We will now evaluate the effectiveness of the proposed segmentation on mammogram classification accuracy. First, Section 4.4 will describe the image features used for classification, i.e. grey level co-occurrence matrix features (GLCM), the database used for the evaluation and the experimental protocol. Results will be discussed in Section 4.5.

4.4 Experiment Setup

To evaluate our proposed techniques based on breast border identification and pectoral muscle removal, we used GLCM features to see how the classification of the normal and abnormal is affected by our proposed segmentation. We will demonstrate our proposed pectoral muscle removal has a positive effect on the classification accuracy of the system.

In the following section, we present the GLCM features that are extracted from the mammograms to classify them as normal or abnormal (Fatty, Glandular, and Dense). Also explained is the Mini-MIAS dataset that was used to evaluate the result and the experimental strategy.

4.4.1 Dataset and Strategy

The proposed method was applied on the complete Mini-MIAS database (Suckling, et al., 1994). In our work, we separate the mammograms into three distinct categories and apply feature extraction and classification on each group individually to separate normal and abnormal mammogram.

The leave-one-out strategy was used to evaluate the accuracy of the system. In this strategy, each time only one mammogram was taken off the database for testing and all the remaining mammograms were used as training. Experiments were repeated (m) times, where (m) is the total number of mammograms in the database which is 322 for Mini-MIAS database, the accuracy rate is averaged. The classification is executed using KNN and SVM as described in Section 4.4.3.

4.4.2 Gray level Co-occurrence Matrix Features

A most commonly used statistical description of intensity values is the intensity histogram, which is the first-order statistical analysis of the image. Some texture features are based on the second-order statistics of image properties, where the relationship between a pair of pixels is analysed to describe the spatial configuration of texture.

Gray Level Co-occurrence Matrix (GLCM) is one of the texture features based on pixel neighbours, and it is feature are widely used in mammogram analysis (Sharma, et al., 2012). The GLCM is used to characterize texture patterns; it can be constructed with different distances (i.e. for a selected pixel, GLCM matrix can be constructed with the first neighbour, $d=1$, or with second neighbour $d=2$ and so on). Also, GLCM matrix can be obtained in different direction angles such as 0, 45, 90, and 135 degrees (Sabu, et al., 2012). Gray level co-occurrence matrix (GLCM) is the basis for the Haralick texture features (Haralick, et al., 1973). Haralick proposed to use a co-occurrence matrix to describe the two-dimensional spatial dependence of grey levels for a fixed distance and/or angular spatial relationship. The GLCM matrix is square with dimension N_g , where N_g is the number of gray levels in the image. Example for extracting GLCM matrix is shown in Figure 4.7.

0	0	4	5	7
1	2	4	6	0
3	4	6	0	1
7	4	0	1	4

Part of original image

	0	1	2	3	4	5	6	7
0	1	2	0	0	1	0	0	0
1	0	0	1	0	1	0	0	0
0	0	0	0	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	1	0	0	0	0	1	2	0
5	0	0	0	0	0	0	0	1
6	2	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0

GLCM matrix

Figure 4.7: Example of extracting GLCM matrix from a part of image

We extracted fourteen texture measures as suggested by Haralick from the GLCM based on the first neighbour and at 0° angle to represent a mammogram image. In the following equations $p(i, j)$ is the (i, j) th entry in normalized co-occurrence matrix; N_g denotes the dimension of co-occurrence matrix (number of gray levels), and $p_x(i)$ and $p_y(j)$ are the marginal probabilities:

$$p_x(i) = \sum_{j=0}^{N_g-1} p(i, j) , \quad p_y(j) = \sum_{i=0}^{N_g-1} p(i, j)$$

The fourteen texture measures are:

- 1- Angular second moment (ASM) feature: ASM measures the uniformity of an image. The smoothness of the image can be extracted by ASM feature. When pixels are very similar, the ASM value will be large.

$$ASM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i, j)^2 \quad (4.1)$$

- 2- Contrast feature: contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbours.

$$Contrast = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i, j)^2 \right\} , \text{ where } n = |i - j| \quad (4.2)$$

- 3- Entropy Feature: this measure analyses the randomness. It is high when the probability (p) values of the selected block have similar values. It is low when the p values are close to either 0 or 1 (i.e. when the pixels in the local block are uniform).

$$Entropy = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j) \log(p(i,j)) \quad (4.3)$$

- 4- Correlation Feature: the correlation feature is a measure of gray-level linear dependency of the image.

$$Correlation = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j) \frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y} \quad (4.4)$$

Where μ and σ are the mean and standard deviation of the co-occurrence matrix respectively:

$$\mu_x = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} i * p(i,j) \quad , \quad \mu_y = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} j * p(i,j) \quad (4.5)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i - \mu_x)^2 * p(i,j)} \quad , \quad \sigma_y = \sqrt{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (j - \mu_y)^2 * p(i,j)} \quad (4.6)$$

- 5- Inverse Difference Moment (IDM) Feature: Inverse difference moment is the measure of local homogeneity of an image.

$$IDM = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{1}{1 + (i - j)^2} p(i,j) \quad (4.7)$$

- 6- Sum Average Feature:

$$Sum\ Average = \sum_{i=0}^{2(Ng-1)} i * p_{x+y}(i) \quad (4.8)$$

Where p_{x+y} is the summation of all $p(i, j)$ in the row and column of the normalized co-occurrence matrix for $k=i+j = \{0, 1, 2, 3 \dots 2(Ng-1)\}$, and calculated as follows:

$$p_{x+y}(k) = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i, j) \quad (4.9)$$

7- Sum Variance Feature:

$$Sum\ Variance = \sum_{i=0}^{2(Ng-1)} (i - Sum\ Entropy)^2 p_{x+y}(i) \quad (4.10)$$

8- Sum Entropy Feature:

$$Sum\ Entropy = - \sum_{i=0}^{2(Ng-1)} p_{x+y}(i) \log p_{x+y}(i) \quad (4.11)$$

9- Difference Variance Feature:

$$Difference\ Variance = \sum_{i=0}^{Ng-1} (i - F)^2 p_{x-y}(i) \quad (4.12)$$

Where p_{x-y} is the summation of all $p(i, j)$ in the row and column of the normalized co-occurrence matrix for $k=i-j = \{0, 1, 2, 3 \dots 2(Ng-1)\}$, and calculated as follows:

$$p_{x-y}(k) = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i, j), k = |i - j| = \{0, 1, 2, \dots, (Ng - 1)\} \quad (4.13)$$

$$F = \sum_{i=0}^{Ng-1} i * p_{x-y}(i) \quad (4.14)$$

10- Sum of Squares :Variance features:

$$Sum\ of\ squares = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (i - \mu)^2 p(i, j) \quad (4.15)$$

11- Information Measures of Correlation1:

$$inf1 = \frac{HXY - HXY1}{\max(HX, HY)} \quad (4.16)$$

Where: HX and HY are entropies of p_x and p_y , and

$$HXY1 = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i, j) \log\{p_x(i)p_y(i)\} \quad (4.17)$$

$$HXY2 = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p_x(i) p_y(j) \log\{p_x(i)p_y(j)\} \quad (4.18)$$

12- Information Measures of Correlation2:

$$inf2 = (1 - \exp[-2(HXY2 - HXY)])^{1/2} \quad (4.19)$$

$$HXY = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i, j) \log(p(i, j))$$

13- Difference Entropy:

$$Difference\ entropy = - \sum_{i=0}^{Ng-1} p_{x-y}(i) \log\{p_{x-y}(i)\} \quad (4.17)$$

14- Maximal Correlation Coefficient:

$$M = (\text{second largest eigen value of } Q)^{1/2} \quad (4.18)$$

Where:

$$Q(i, j) = \sum_k \frac{p(i, k)p(j, k)}{p_x(i)p_y(j)} \quad (4.19)$$

4.4.3 Classification Methods

The process of differentiating two or more classes is called classification. In any CAD system, classification is the final and main task which aims to identify cases that are truly recognised as exhibiting the sought after abnormality and distinguish them from the cases which are not recognized. Examples of classifiers include Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees and K-nearest neighbour (KNN) method. A typical classification system work in two phases: the training phase and the testing phase. Training is based on using a sample data obtained from available records and aim to learn discriminating factors about the extracted features and determine digital template representative of the different classes to be compared later in order to identify or verify the class of fresh samples. The testing phase is where comparison scores between test samples, separate from the training set, with the saved templates in order to determine the class that this set of data should belong to. The testing phase determines the accuracy of the algorithm with respect to the given classifier. In this chapter, we use two different classifiers, namely K-Nearest Neighbour (KNN) (Guo, et al., 2003) and Support Vector Machines (SVM) (Gunn, 1998).

4.4.3.1 K-Nearest Neighbour (KNN)

The K-Nearest Neighbour is a kind of lazy classifier, meaning there is no learning or training for a model using known data to classify known data. KNN are very expensive when it is applied to a large data set. KNN is suitable for multi-modal classes as its classification decision is based on a less distance neighbourhood of similar objects, and a tie is broken randomly. The details of KNN can be found in (Gonzales & Woods, 2002) (Guo, et al., 2003). In our work K=1 and a Euclidean distance function is used. If you have two feature vectors u and v , the Euclidean distance measurement is:

$$D_{Euclidean}(u, v) = \sqrt{\sum_i (u_i - v_i)^2} \quad (4.20)$$

4.4.3.2 Supported Vector Machine (SVM)

Support vector machines (SVM) is a very good and powerful technique for classification purposes. SVM has been applied to different problems such as face recognition (Heisele, et al., 2001), cancer diagnosis and text categorization. SVM divides the given data into decision surface. Decision surface divides the data into two classes like a hyper plane. Training points are the supporting vectors which define the hyper plane. In SVM classifier, the basic property is to maximize the margins between two classes of the hyper plane (see Figure 4.8). However, the main advantage of SVM is reducing dimensionality when the feature vector is very large. In this chapter, we trained a binary SVM classifier to classify a given mammogram image to fatty, glandular or dense as a normal or abnormal mammogram. SVM classifier is widely used in mammogram images (Tzikopoulos, et al., 2010). The details of SVM can be found in (Gunn, 1998) (Duda, et al., November 2000).

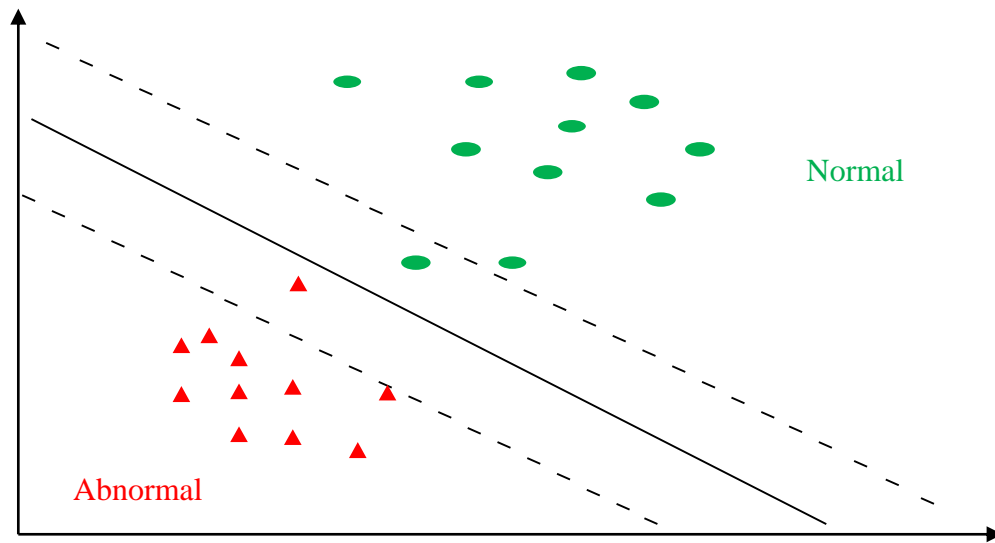


Figure 4.8: An example of SVM model

4.5 Results and Discussion

The proposed breast boundary detection technique, which is based on a simple inference, gives satisfactory results. This is clear by a careful observation of the detected boundary of the mammogram images. The two proposed methods (breast border extraction and pectoral muscle detection) were evaluated with the help of an expert radiologist. For the first proposed breast border extraction method, 112 cases

chosen randomly from the Mini-MIAS database, were voted by an expert radiologist, the decision for the expert was that the borders were extracted successfully for all 112 cases (i.e. 100% cases accepted), and for the pectoral muscle removal method, 94 cases chosen randomly, were voted by the expert radiologist and found 62 cases to be ranked as acceptable, 19 cases nearly acceptable, and 13 cases as not acceptable (i.e. 86.17% are accepted or nearly accepted). Pectoral muscle removal is considered unacceptable by radiologists if part or all the muscle remained after applying the specified technique.

Comparing the results we obtained here with those results obtained from the previous studies that we reviewed in Section 3.2, where similar visual evaluation criteria were used, we can see that our proposed method achieved better results. For example, Chen and Zwiggelaar (Chen & Zwiggelaar, 2010) achieved 98.4% of acceptable segmentation (border extraction) on the test set of 240 mammograms from the EPIC database of which 66.5% were accurate, 25% nearly accurate and 6.9% acceptably segmented. While for the pectoral muscle removal method they achieved 87.9 % nearly accurate. Mustra (Mustra & Grgic, 2013) used Mini-MIAS database and proposed schemes obtained 99.06 % of acceptable border extraction segmentation and about 89.69 % of successful pectoral muscle removal when the proposed method was tested by an expert radiologist.

Moreover, the novelty of our segmentation based method, is its simplicity and effectiveness in comparison to the different above mentioned approaches. Our method simply traces the visible thin line separating the breast tissue from the pectoral muscle. This visibility of the thin line is due to noticeable differences between pixel intensities on both sides of the pectoral muscles border. Samples for border extraction and muscles removal with different decisions made by radiologists shown in Figure 4.9 and Figure 4.10, respectively.

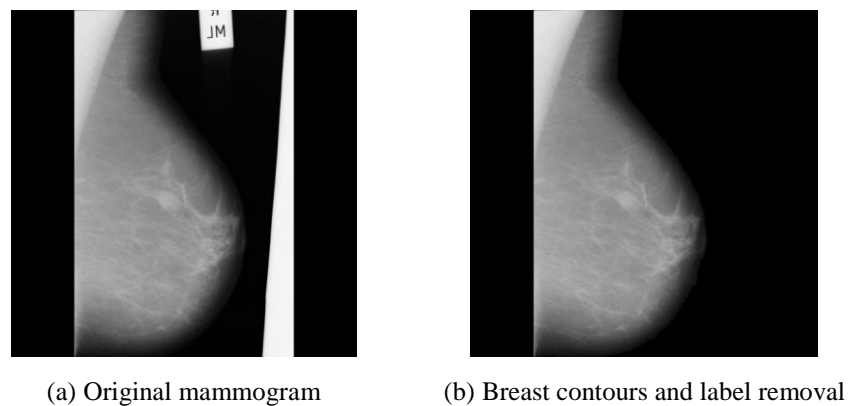


Figure 4.9: Label removal and Border extraction sample

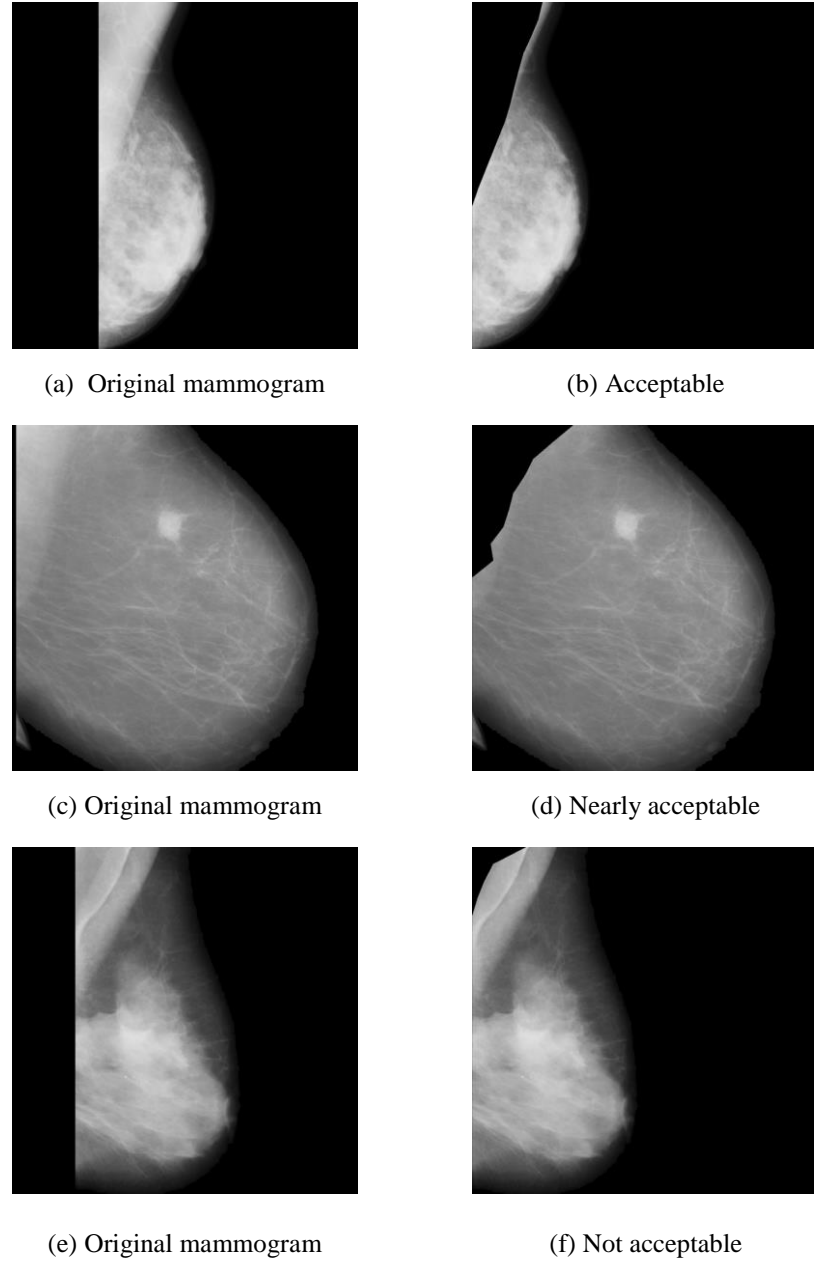


Figure 4.10: Radiologist Decisions for muscle removal

Furthermore, we evaluated our proposed segmentation method by calculating the sensitivity, specificity and accuracy for each class of mammogram images in the database by leave-one-out strategy for the original mammogram images and for the segmented mammogram images. Table 4.1 to Table 4.4 shows the result of the classification for three categories of mammograms. For each case, sensitivity, specificity and accuracy rates of the system are calculated. **Note that, no explicit consideration were given to sensitivity and specificity in existing schemes.**

Table 4.1: Classification result, $d=1$, angle=0 degree

Type	Measure	Original Image		Proposed method	
		SVM	KNN	SVM	KNN
Fatty	Sensitivity	10.25	15.38	82.05	89.74
	Specificity	100	58.20	98.50	97.01
	Accuracy	66.98	42.45	92.45	94.33
Glandular	Sensitivity	67.56	48.64	40.54	37.83
	Specificity	82.08	53.73	100	92.53
	Accuracy	76.92	51.92	78.84	73.07
Dense	Sensitivity	5.5	16.66	27.77	33.33
	Specificity	100	69.73	97.36	86.84
	Accuracy	69.64	52.67	75	69.64

Table 4.2: Classification result, d=1, angle=45 degree

Type	Measure	Original Image		Proposed method	
		SVM	KNN	SVM	KNN
Fatty	Sensitivity	10.25	28.20	82.05	84.61
	Specificity	100	65.67	98.50	98.50
	Accuracy	66.98	51.88	92.45	93.39
Glandular	Sensitivity	21.62	37.83	48.64	43.24
	Specificity	85.07	58.20	100	85.07
	Accuracy	62.5	50.96	81.73	70.19
Dense	Sensitivity	5.55	22.22	47.22	58.33
	Specificity	98.68	73.68	92.10	82.89
	Accuracy	68.75	57.14	77.67	75

Table 4.3: Classification result, d=1, angle=90 degree

Type	Measure	Original Image		Proposed method	
		SVM	KNN	SVM	KNN
Fatty	Sensitivity	7.69	28.20	94.87	89.74
	Specificity	98.5	76.11	98.50	92.53
	Accuracy	65.09	58.49	97.16	91.50
Glandular	Sensitivity	13.51	27.02	45.94	45.94
	Specificity	91.04	67.16	97.01	89.55

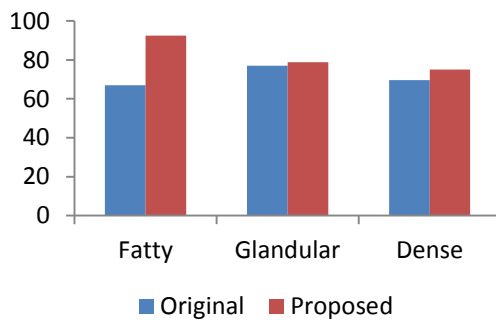
Type	Measure	Original Image		Proposed method	
		SVM	KNN	SVM	KNN
	Accuracy	63.46	52.88	78.84	74.03
Dense	Sensitivity	13.88	27.77	8.33	30.55
	Specificity	98.68	78.94	98.68	82.89
	Accuracy	71.42	62.5	69.64	66.07

Table 4.4: Classification result, d=1, angle=135 degree

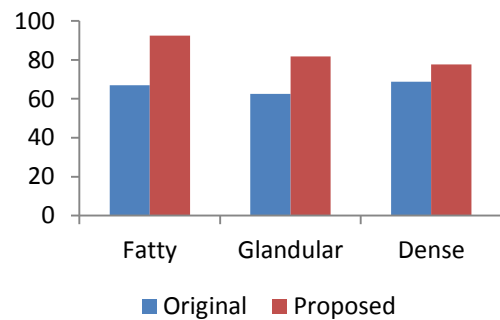
Type	Measure	Original Image		Proposed method	
		SVM	KNN	SVM	KNN
Fatty	Sensitivity	7.69	23.07	82.05	74.35
	Specificity	100	67.16	98.50	95.52
	Accuracy	66.03	50.94	92.45	87.73
Glandular	Sensitivity	24.32	27.02	45.94	51.35
	Specificity	79.10	47.76	100	86.56
	Accuracy	59.61	40.38	80.76	74.03
Dense	Sensitivity	5.55	33.33	30.55	22.22
	Specificity	100	68.42	98.68	88.15
	Accuracy	69.64	57.14	76.78	66.96

From the tables, it is clear that deleting the pectoral muscle and artifacts using our algorithms have improved the classification accuracy in most cases. Furthermore, if we look at the results obtained from using original mammograms for extracting features we can see unbalance between results of sensitivity and specificity, obtained specificity are much higher than sensitivity (i.e. abnormal cases are recognised wrongly as normal). We noticed an important thing, which is that the extracted features are more accurate for the fatty mammogram than the other two types (Glandular and dense) mammogram. A similar observation was made by (Elshinawy, 2010) where it is mentioned that the texture features extracted using GLCM are more sensitive to the changes of the intensity values of the fatty mammogram than the other types. If we look at the results of glandular and dense images for the original mammogram, we notice that there is an imbalance between sensitivity and specificity of the system because the label and the muscles affect the results. For specific comparison between using KNN and SVM classifier, results show that in most cases using SVM results in better accuracy than KNN. Extracting GLCM from different directions also have an

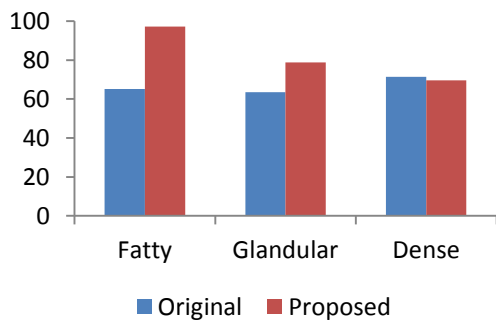
effect on the results; we can say that extracting GLCM features using directions 0 degree and 90 degree have better accuracy than the others. Figure 4.11 and Figure 4.12 show the diagram of the classification accuracy by using KNN and SVM classifiers respectively. For example, if we look to Figure 4.11 (b), the accuracy results for the original mammogram using KNN classifier are 66.98%, 62.5%, and 68.75% respectively for fatty, Glandular and, dense mammograms. We can see that the classification accuracy for the system is increased after deleting the pectoral muscles and artifacts in the mammogram images (i.e. applying the proposed method). So, the classification accuracy for fatty mammograms is 92.45%, for glandular mammograms, it is 81.73% and for dense mammograms, it is 77.67% after segmenting using the proposed methods. Also for the same case in Figure 4.12(b), the accuracy for the original mammogram using SVM classifier are 51.88%, 50.96% and 57.14% respectively for fatty, glandular and, dense mammogram images. The classification accuracy increased after removing the pectoral muscle and artifacts (i.e. applying the proposed method). After segmenting using the proposed methods, the classification accuracy for fatty mammograms is 93.39%, for glandular mammograms, it is 70.19% and for dense mammograms, it is 75%. The two evaluations demonstrate the effectiveness of the proposed pre-processing techniques to extract breast borders and pectoral muscle removal from the mammogram.



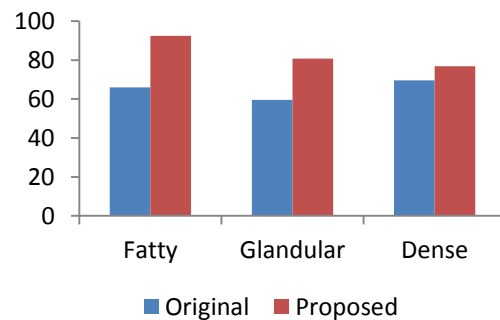
(a)



(b)



(c)



(d)

Figure 4.11: Classification accuracy by KNN classifier, (a) $d=1$ and angle=0 degree, (b) $d=1$ and angle= 45 degree, (c) $d=1$ and angle= 90 degree and (d) $d=1$ and angle=135 degree

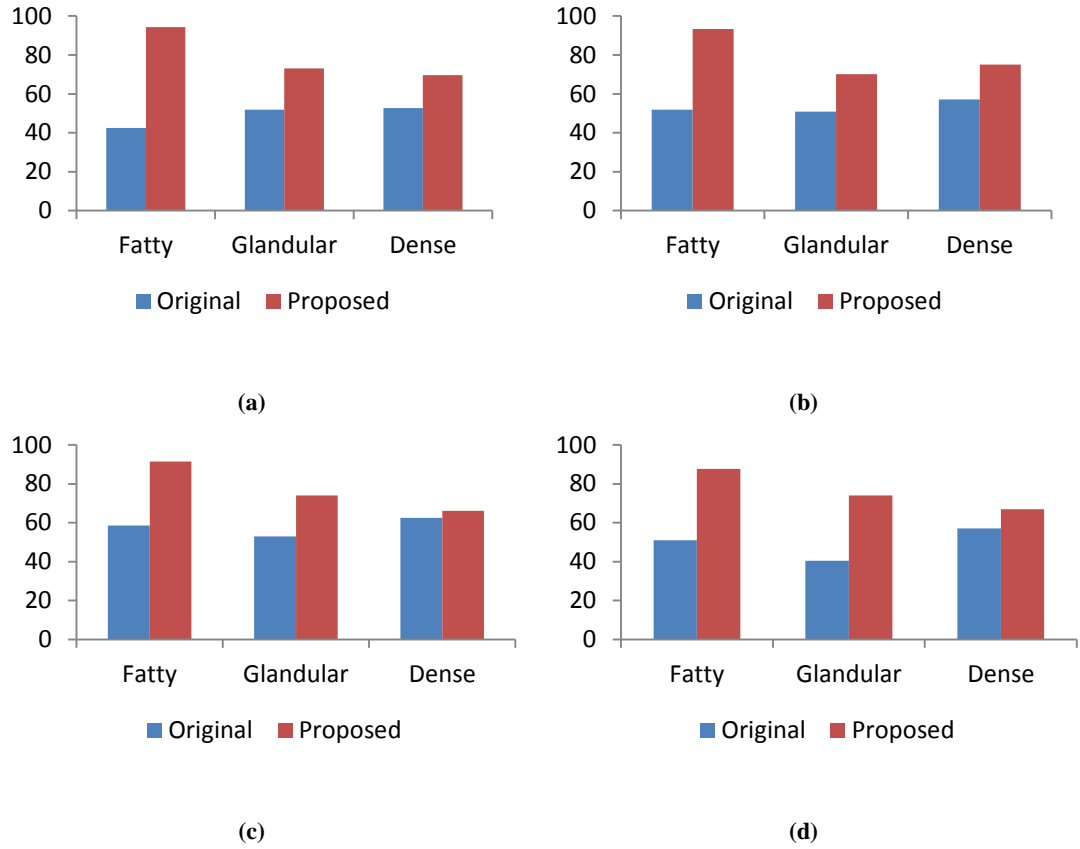


Figure 4.12: Classification accuracy by SVM classifier, (a) $d=1$ and angle=0 degree, (b) $d=1$ and angle= 45 degree, (c) $d=1$ and angle= 90 degree and (d) $d=1$ and angle=135 degree

4.6 Conclusions

Segmentation is a very important step in any CAD system. The extraction of the breast contour is useful because it limits the search-zone for abnormalities. Moreover, in most Medio-Lateral Oblique (MLO) views the pectoral muscle represents a predominant density region in the breast tissue. To avoid the bias of the detection procedure in a CAD system we need to identify and segment out the pectoral muscle during pre-processing stage. Designing an automated algorithm capable of segmenting the breast region in mammograms and deleting the pectoral muscles is a difficult task. In this chapter, we proposed a new method for breast border extraction, artifact removal and removal of annotations. Furthermore, we proposed an adaptive algorithm to detect and delete the pectoral muscle automatically; the algorithm is based on the gray-level intensity values

From the experimental results shown in this chapter, we conclude that using mammograms after accurately removing the unwanted parts such as labels, artifacts, and pectoral muscle yield better classification accuracy than using the entire mammograms. This is because the pectoral muscle region as well as labels and artifacts in mammograms have pixel intensity values that are similar to the pixel intensity of the abnormal parts inside the breast area and thereby detecting abnormalities become less accurate.

Due to the different size and shapes of the breast regions, the remaining part of the mammogram after removing unwanted parts could be different in sizes and shapes. These differences may affect the classification accuracy of the system or affect the ability to use them for specific kind of features extraction methods that require fixed size of images. Therefore, in the remaining chapters, we shall use cropped fixed size mammograms for the purpose of evaluating the performance of the proposed schemes.

Chapter 5

Texture-based Features for Mammogram Classification

One of the most important processes in any classification system, including mammogram classification, is the feature extraction process. The extracted features from individual mammogram images represent that mammogram. The way of extracting features from mammograms are very important as they affect the accuracy of the mammogram classification system. In this chapter, we shall investigate the use of different texture feature extraction schemes to design a new, efficient and accurate schemes for mammogram classification. We investigate the use of multi-scale local binary patterns where local binary pattern histograms (LBPH) are extracted from wavelet sub-bands of mammogram images. Different combinations of LBPH over wavelet sub-bands will be used to increase mammogram classification accuracy. LBPH features extracted in the spatial domain as well in the wavelet domain will be combined to make a feature vector to represent mammogram images. We also investigate the use of two other texture features, which are GLCM and HOG in mammography field.

Finally, we investigate the use of PCA as a dimension reduction scheme to reduce the feature vectors sizes/dimensions that represent mammogram images (or their ROIs).

Section 5.1 describes the different texture features (i.e. LBP, DWT, HOG and GLCM) considered in this thesis. Our proposed approaches to extract multi-scale texture features from mammograms are explained in Section 5.2. Then Section 5.3 describes the experiments and the evaluation scheme designed to test the performance of the proposed feature extraction schemes. Experimental results will be analysed and discussed in Section 5.4. A summary of the results is presented in Section 5.5. Sections 5.6 and 5.7 will explain the concept of dimension reduction as well as the investigation of applying dimension reduction schemes over the feature vectors. Our concluding remarks for this chapter are offered in Section 5.8.

5.1 Background

Generally, image features are classified into four main types; statistical features (such as mean, standard deviation and smoothness), textural features (such as local binary pattern, gray level co-occurrence matrix, and histogram of oriented gradient), model based features (such as features based on Markov random field and fractal models), and signal processing features (such as Gabor features, wavelets, Fourier transform) (Elshinawy, 2010).

From the literature we can conclude that; there is no specific single set of features known to produce optimum results for mammogram classification (i.e. there is no ability to find the set of features that best suits each type of tissue for mammograms). The main reason behind this is the irregularity and randomness of the abnormal mammogram patterns. This issue remains the main challenge in mammogram feature extraction process.

In the field of mammogram classification, much attention has been given to texture features based classification (Ponraj, et al., 2011). Therefore, here in this thesis, we used different types of texture feature extraction schemes such as Local Binary Patterns (LBP), wavelet-based LBP, Histogram of Oriented Gradients (HOG), and Gray Level co-occurrence matrix (GLCM). Each of these texture features has their own specific, and complementary, characteristics and properties, which we will explain in the corresponding subsections below.

5.1.1 Discrete Wavelet Transform (DWT)

A Wavelet Transform (WT) can be described as a mathematical tool that produces components (coefficients) of different frequencies of a specific signal (i.e. is a process of representing the signal in different scales) (Gonzales & Woods, 2002) (Sellaheewa, 2006).

Coefficients in the WT domain have the properties of localization and correlation to the spatial domain (Mallat, 1989). The correlation of the wavelet coefficients means that each coefficient can be mapped from the wavelet domain to the spatial domain (Gonzales & Woods, 2002) (Al-Jawad, 2009). This property makes WT coefficients become a commonly used tool in a very wide range of applications such as image compression, biometric feature extraction, object detection and recognition, image and

video content based retrieval, information hiding, image de-noising, image enhancements and more (Gonzales & Woods, 2002).

When applied to an image, the DWT produces four different sub-bands (LL, HL, LH, HH), with each containing different information of the original image. Each sub-band can be decomposed further to achieve a multi-resolution analysis of the original image. Two example images and the results of their wavelet decomposition at two levels are illustrated in Figure 5.1.

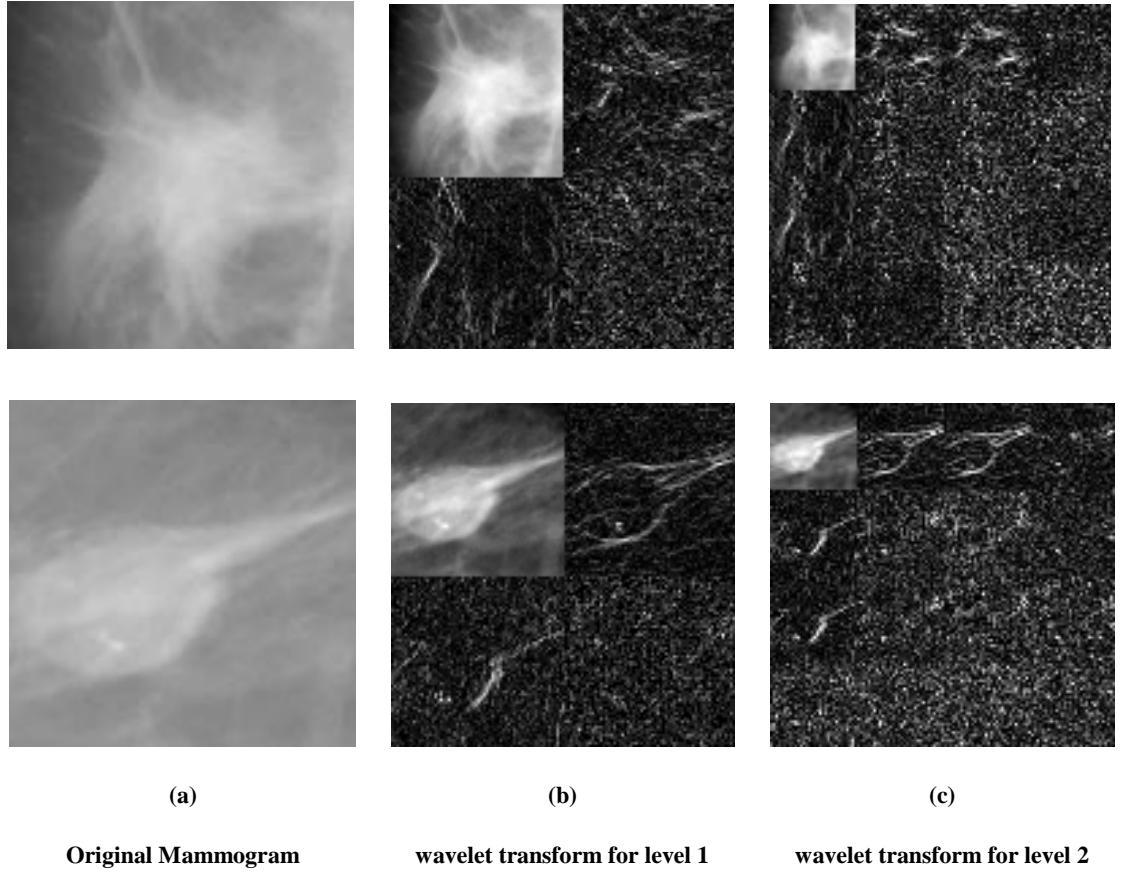


Figure 5.1: Wavelet Decompositions

The most common decomposition technique is the pyramid decomposition where only the LL sub-band is used for the decomposition.

For the pyramid decomposition scheme, the LL sub-bands at each resolution level is further decomposed to four sub-bands. Therefore, an image I is decomposed into $3k+1$ sub-bands ($LL_k, HL_k, LH_k, HH_k, \dots, HL_1, LH_1, HH_1$), with LL_k being the lowest-pass sub-band. The sub-bands LH_1, HL_1 , and HH_1 contain the finest scale wavelet coefficients. The LL_k sub-band is considered as the k^{th} -level approximation of the original image, I ; LH and HL sub-bands represents horizontal and vertical features of

the image respectively, while HH sub-bands represent the diagonal features of the image.

Results of pyramid wavelet decomposition at three levels of two mammogram images are illustrated in Figure 5.2.

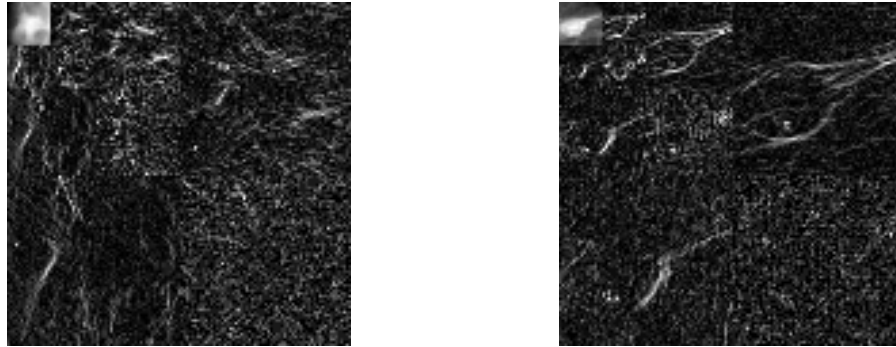


Figure 5.2: Three Level Pyramid DWT decomposition of mammograms

Previous studies reported in the literature have shown that each individual wavelet sub-band (low frequency LL and high-frequency wavelet sub-bands LH, HL, and HH) contain useful, and often complementary, information about the original spatial image that can be used as a feature set to represent the image (or the object of interest in the image) for the purpose of classification. Therefore, several studies have proposed the fusion of features extracted from the different wavelet sub-bands to represent the image/object of interest. For example, Sellahewa & Jassim (Sellahewa & Jassim, 2010) (Sellahewa & Jassim, 2008) fused multiple sub-bands deal with the problem of varying illumination and expressions in face recognition. We too use the idea of multi wavelet sub-band fusion to represent mammogram images to capture texture/features at different scales and orientations.

5.1.2 Local Binary Patterns (LBP) Features

Proposed by Ojala et al., (Ojala, et al., 1996), the LBP operator is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local pixel neighbourhood. There are number of advantages of using LBP operator such as: (i) extracting LBP features has a very low computational complexity, which makes real-time image analysis possible, (ii) LBP features are invariant to monotonic gray level changes, and (iii) LBP features have excellent discriminative power which makes them suitable for classification problems (Zhao & Pietikainen, 2007). LBP features are widely used in detection and recognition in other fields of image processing such as the

face (Ahonen, et al., 2004) and palmprint (Guo, et al., 2010) recognition, LBP also used in mammogram field as described and discussed in Chapter 3 Section 3.3 of this thesis.

The LBP code is obtained by replacing selected pixel value with 8-bit binary code which is derived from its 8 neighbour pixel values. The process of obtaining these codes starts by subtracting the value of the centre pixel from each of 8 neighbour pixel values then depends on the sign of the subtraction result, 1 or 0 assigned to a bit location. The generated 8 bits for all the 8 neighbours pixels are concatenated and encoded into binary strings in a clockwise direction. The final 8 binary string are called LBPs codes see Figure 5.3. The decimal value of the LBP code for the centre pixel (x_c, y_c) is calculated as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{n-1} s(i_n - i_c) 2^n \quad (5.1)$$

i_c and i_n are gray level values of the central pixel and the surrounding pixels respectively, and the function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (5.2)$$

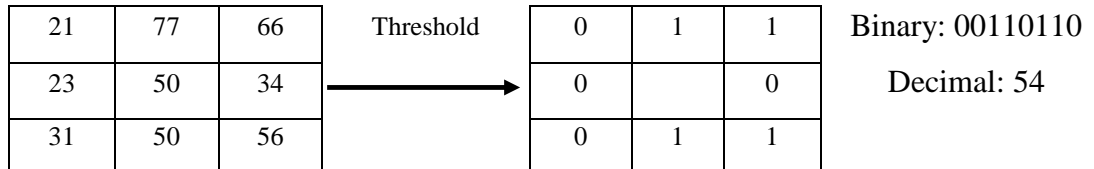


Figure 5.3: An example of LBP operator

After an image is labeled with the LBP operator, a histogram of the labeled image is calculated. This LBP histogram used as a texture feature that contains information about the distribution of the local micro-patterns, such as edges, spots, and flat areas, over the whole image. Symbolically, the LBP operator can be represented as $LBP_{P,R}$ where 'P' represents a number of neighbours used to calculate LBP code and 'R' is the radius which mentions how neighbours are far from the centre pixel. $LBP_{P,R}$ produces 2^P different output values, corresponding to 2^P different binary patterns formed by the P pixels in the neighbourhood of radius R for example $LBP_{8,1}$ produce 256 different output values formed by 8 first neighbours.

Again (Ojala, et al., 1996) demonstrated that certain patterns contain more information than others and that it is possible to use only a subset of the 2^p binary patterns to represent texture. These patterns were named uniform patterns, denoted by $LBP_{p,R}^{u2}$. The uniform patterns are those patterns which contain at most two bitwise transitions from 0 to 1 or from 1 to 0 when the corresponding bit string is considered circular. For instance, 00000000 (0 transitions) and 11101111 (2 transitions) are uniform whereas 11110110 (4 transitions) and 10110110 (6 transitions) are not uniform patterns.

From experiments, Ojala et al. showed that 90% of all patterns in case of using $LBP_{8,1}$ are applying uniform patterns when texture images are used. In case of using $LBP_{8,1}$, from 256 patterns there are only 58 uniform LBP-patterns, and all other 198 are non-uniform patterns. Therefore, the histogram that represents this kind of features consists of 59 bins representing 58 uniform patterns and 1 bin represents the sum of all 198 remain non-uniform patterns. Using this technique to extract features will greatly reduce the number of features that represent any texture images.

Ahonen et al. (Ahonen, et al., 2004), showed that the histogram of LBP computed over the whole face image encodes only the occurrences of the LBP patterns without giving any hint regarding their locations. Therefore, to consider the aspect of ‘shape information’ of faces, authors proposed the idea of blocking before calculating LBP code of the image. They first divided the face image into m local regions to extract LBP histograms (LBPHs) from each region and concatenated them into a single feature histogram to represent the whole face image. Figure 5.4 illustrates the idea of extracting region or block based LBP. Most of the existing works, including our proposed used of LBP, adopt the above scheme to extract LBP features for image representation.

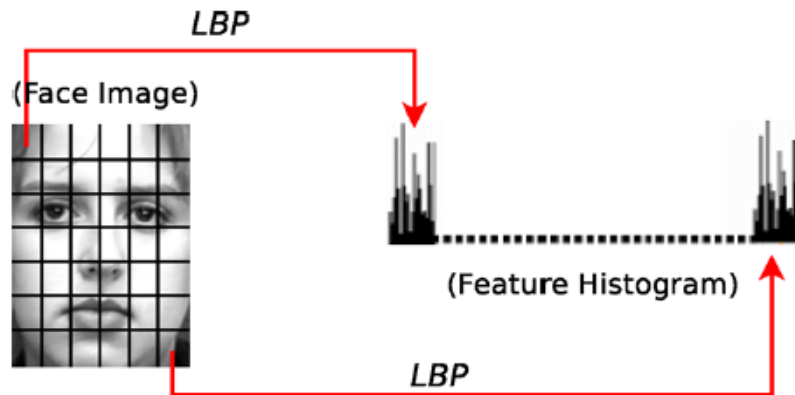


Figure 5.4: Sub block histogram concatenating (Ahonen, et al., 2004)

5.1.3 Multi-scale LBP

The basic LBP operator has one limitation which is its small 3×3 neighbourhood cannot capture dominant features with large-scale structures. To overcome this limitation and calculate texture at different scales, the LBP operator was later extended (Ojala, et al., 2002) and generalized to use neighbourhoods of different sizes. A local neighbourhood is defined as a set of sampling points spaced on a circle, which is centred at the pixel to be labeled, thus allowing for any radius and any number of sampling points in the neighbourhood. Figure 5.5 shows some examples of the extended LBP operator.

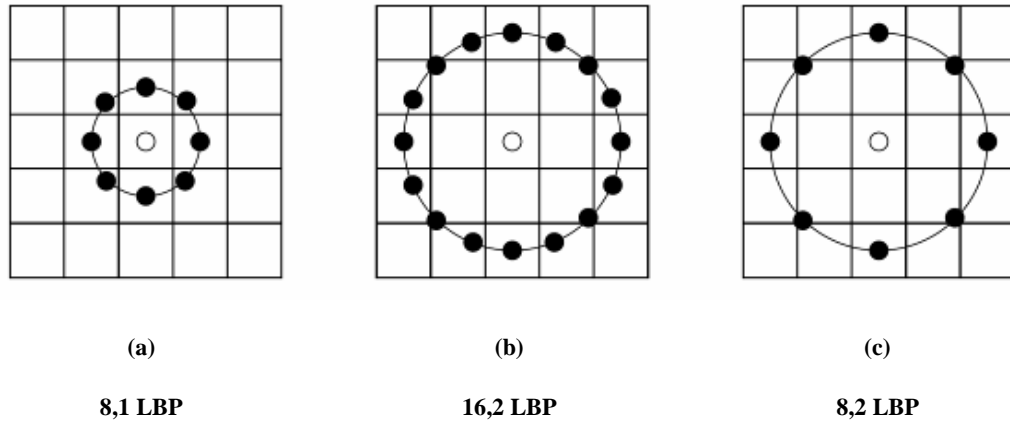


Figure 5.5: Circular neighbourhoods for LBP (a), (b), and (c)

There are two ways of calculating multi-scale LBP histograms; by calculating LBP code of the centre pixels with different R and then concatenate all histograms into a large histogram representation, or by applying the wavelet transform over the original image prior to extracting the LBP code with the same radius (R) of different scales, and then concatenate all histograms together to represent final feature vector (Y. Wang, 2011) (Rashid, et al., 2013). Here in this thesis when we use multi-scale LBP features, and we use the second approach which is to apply a wavelet transform first followed by calculating LBP code on wavelet sub-bands.

5.1.4 Histogram of Oriented Gradients (HOG) descriptor

Local object (inside an image including mammograms) appearance, as well as object shape, can be characterized by the distribution of local intensity gradients or edge direction. The Histogram of Oriented Gradients (HOG) descriptor depends on the distribution of a differential intensity histogram of an image. In the HOG descriptor,

the image is blocked by non-overlapping uniform cells. For each cell, the intensity variation in different directions is calculated (Dalal & Triggs, 2005).

First described by Dalal & Triggs (Dalal & Triggs, 2005), they focused on the use of HOG features in the problem of pedestrian detection in static images. Also, they expanded their tests to include human detection in film and video streams, as well as to a variety of common animals and vehicles in static imagery. Subsequently, many studies used HOG descriptor for the purpose of human detection such as in (Kobayashi, et al., 2008) to classify the given input into pedestrian/non-pedestrian. The HOG descriptor has been successfully used in different computer vision areas such as face detection and recognition (Déniz, et al., 2011).

The local shapes of an object that include information about edges in a given cell are well represented at the final HOG feature vector. Moreover, the HOG has a flatter distribution in the smooth regions of an image (Dalal & Triggs, 2005). For example, mammograms without having mass or any abnormality. While on the border between an object and the background, one of the elements in the histogram has a large value and will indicate the direction of the edge. So researchers in the field of classifying mammograms are now starting to use this kind of features.

Recently, Kirshnaveni et al. (Krishnaveni, et al., 2014) proposed a method to detect microcalcifications in mammograms. The ROI was extracted after that pre-processing and enhancement is applied to remove noise and artifact. Finally, HOG is applied, and GLCM features and intensity features are measured. The mammograms were classified using a Naive Bayes Classifier. Our contribution in this direction is to propose and investigate the use of HOG descriptor for classifying mammograms to normal and abnormal cases rather than detecting microcalcification as in (Krishnaveni, et al., 2014). We use the HOG descriptor in different ways: (1) as a single feature to represent a mammogram; (2) Applying PCA to the HOG feature vector to reduce feature vector size (see Chapter 6); and (3) fuse HOG features with other kind of texture features (e.g. fuse of LBP and HOG features) to represent final mammogram feature vector (see Chapter 6).

5.1.5 Gray level co-occurrence matrix (GLCM)

GLCM are created by taking pairs of image cells a distance 'd' from each other and incrementing the position corresponding to the gray level of both cells inside the

matrix. This will lead to four different matrices one for each given distance: $P(0^0; d)$, $P(45^0; d)$, $P(90^0; d)$, and $P(135^0; d)$. More details were given in Chapter 4 Section 4.4.2. In our work we extract GLCM using four angles (0, 45, 90, 135) degree with distance $d=1$ after that we extract statistical features from GLCM table (the same way of extracting the feature from the mammogram in Chapter 4) but in this chapter we extract GLCM features from the cropped Mini-MIAS and DDSM databases.

5.2 Proposed Feature Extraction Methods

When extracting features from mammogram images, there is no specific technique that works well with all kinds of mammogram images. In this stage of our work, we used and proposed different techniques, and we used each of them separately to extract texture features; the aim is to compare and discuss different kind of texture features regarding the efficiency of classification, by doing so we can conclude that which set of features work better than the other to classify mammogram images. More especially the main contribution in this step is to extract texture features and classify mammogram images to normal and abnormal images.

5.2.1 Multi-Scale LBP

By taking the advantages of using wavelet decomposition to extract multi-scale LBP code, recently, LBPH feature are extracted in the wavelet domain to extract multi-resolution features for face recognition (Tang, et al., 2010) and hand vein recognition (Y. Wang, 2011). Inspired by the successful use of LBP in the wavelet domain for biometric authentication, we first decompose mammogram images into two level using DWT with a Haar filter, secondly we sub-divide the sub-band to number of blocks, here we used 3x3 blocking, thirdly we calculate the LBP code on each sub-block of the wavelet sub-band separately and calculate the histogram of the extracted LBP code. Finally we concatenate all histograms calculated from each sub-block to represent final mammogram image. In all cases for extracting LBP code in each sub-block, we used the LBP operator with 8-neighbours at a radius of 1. Furthermore, we used the uniform LBP (i.e. each histogram represented by only 59 bins) due to its performance advantage and feature/histogram size. Here we used different representations of feature sets based on sub-band or number of sub-bands used to represent final mammogram image such as:

- **Single Sub-band:** In this case, we used only one wavelet sub-band to represent mammogram features. We apply the idea for different sub-bands and for two level wavelet decompositions to see how using different sub-bands can affect the classification accuracy of the system.
- **Multi Sub-bands:** Any derived mammogram block regions or sub-bands in frequency domain provides different discriminative contribution which each of them plays a different role in mammogram classification. Information from a single block or sub-band can represent limited discriminative contribution. So mammogram classification based on single mammogram information cannot achieve satisfactory performance. It is reasonable to fuse more than one type of mammogram information to address the classification task. Since different sub-bands contain different features of the original image, therefore, in this case, we used more than one wavelet sub-bands to represent final mammogram. Different combinations are used, and different numbers of sub-bands are used to see how a number of sub-bands may affect the classification accuracy.
- **Original with sub-band(s):** In this case, we fuse mammogram information from the spatial domain with mammogram information from the wavelet domain. Here we fused the LBPH features obtained from the original mammogram image with LBPH features obtained from the wavelet sub-bands. Different combination used with a different number of sub-bands to see how the classification accuracy affected after combining them with original mammogram features.

5.2.2 GLCM

In this work, we obtain the Gray Level Co-occurrence Matrix (GLCM) of the mammogram which is the spatial distribution of pixel values in an image in four different directions at distance $d=1$. The four different directions that suggested by Haralick 1973 (Haralick, et al., 1973) are 0, 45, 90, and 135. The co-occurrence matrices constructed using the four directions are $p(i; j; d; 0)$, $p(i; j; d; 45)$, $p(i; j; d; 90)$, and $p(i; j; d; 135)$. Then we extract the 14 statistical texture features from the four GLCM matrices individually. Details of all 14 statistical features can be found in (Section 4.4) of this thesis. Different directions are used to see and analysis how each direction can affect the mammogram accuracy.

5.2.3 HOG Descriptor

To extract HOG features from mammogram images we used the same approach to feature extraction that used in (Dalal & Triggs, 2005), the aim is to see how this kind of features can be used to detect abnormalities in mammogram images and then used for classification purpose. In the proposed method, the HOG descriptor extracted by using the following steps:

Step 1: After the mammogram image is read, we find image gradient magnitudes over the whole image pixels. The gradient image is calculated by filtering the image using $[-1, 0, 1]$ and $[-1, 0, 1]^T$ filters for horizontal, and vertical direction respectively see equation (5.3) and (5.4).

$$G_x = \frac{\partial f(x, y)}{\partial x} = \frac{f(x + 1, y) - f(x - 1, y)}{(x + 1) - (x - 1)} \quad (5.3)$$

$$G_y = \frac{\partial f(x, y)}{\partial y} = \frac{f(x, y + 1) - f(x, y - 1)}{(y + 1) - (y - 1)} \quad (5.4)$$

Step 2: Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. In the following equation (5. 5) and (5. 6) we calculate the weighting ($\Delta f(x, y)$) and the orientation ($\theta(x, y)$) for an image pixel:

$$\Delta f(x, y) = \sqrt{(G_x^2 + G_y^2)} \quad (5.5)$$

$$\theta(x, y) = \tan^{-1}(G_x^2/G_y^2) \quad (5.6)$$

Step 3: We compute histogram on cells of typically 8x8 pixels (i.e. 16x16 cells in total). The histogram size is 9 bins.

Step 4: The resulting histograms are normalized (often simply to unit length) with overlapping blocks (50% overlap) of cells (typically 2x2, i.e. 15x15 blocks in total),

Step 5: Finally concatenate all obtained histograms into one vector to represent final feature vector.

The final extracted feature vector size using this scheme is 8100 features which result from the following simple calculation:

$$\begin{aligned}\text{Number of features} &= \text{No. of Blocks} \times \text{Block Size} \times \text{No. of Bins} \\ &= (15 \times 15) \times (2 \times 2) \times 9 \\ &= 8100\end{aligned}$$

5.3 Experimental Setup

This section describes the experimental setup used to test the performance of the proposed feature extraction schemes. More specifically, we describe the experimental databases, classification methods, and the evaluation protocols that govern the partitioning of the data samples into training and testing sets.

5.3.1 Databases

We used two different databases to evaluate our proposed feature extraction schemes. The two databases are commonly used in the researches, and they have different properties and structures. Following sub-sections are details of the way that the individual databases are used for the classification purpose.

5.3.1.1 *Mini-MIAS database*

The proposed feature extraction schemes were tested on the complete set of 322 mammograms of the Mini-MIAS database (Suckling, et al., 1994). This database is widely used to evaluate automated mammogram analysis techniques, and it is available freely for scientific research purposes. The database consists of 161 pairs of Medio-lateral oblique (MLO) view mammograms (right and left view) and overall includes 209 normal cases and 113 abnormal cases. The images of the database originate from a film-screen mammographic imaging process in the United Kingdom National Breast Screening Program. For more details, see (Chapter 2 Sub-Section 2.4.1). The Mini-MIAS database contains different types of the abnormality. We separate normal mammogram and abnormal mammogram with different types of the abnormality. We extracted the region-of-interest (ROI) of size 128×128 pixel from the original mammogram image based on the given (x, y) image-coordinates that represent the centre of abnormality when normal images are used, we selected the (x, y) centre of the breast as the image co-ordinate. The features described in previous sections were extracted only from the selected ROI.

5.3.1.2 DDSM database

The Digital Database for Screening Mammography (DDSM) is commonly used as a benchmark for testing new proposals dealing with processing and analysis of mammograms for breast cancer detection (Heath, et al., 2001). Experiments were performed on 512 mammogram images selected randomly out of 2620 images. Of these 512 images, 256 images are normal, and other 256 images are abnormal. All images were cropped and selected ROI of size 128 x 128 were used for experiments. See (Chapter 2, Sub-Section 2.4.2) for more details. Again features described in previous sections were extracted only from the selected ROI.

5.3.2 Evaluation Protocol

We conducted experiments in four different scenarios to evaluate our proposed feature extraction schemes. The four scenarios are different to each other in the way they separate the mammogram database in to training and testing sets (i.e. the number of images used for training and testing are different). Testing the proposed schemes under different scenarios is important, because in real life CAD system, different number of mammogram cases are available (i.e. sometimes fewer number of cases are available while in other larger number of cases are available). We assigned a different percentage of mammograms to the training set (and testing set) to see effects of the training set size on the classification accuracy of the system. The four different scenarios are:

- **Leave-One-Out strategy:** in this strategy, one mammogram is kept as a test image and all the remaining images were used as training samples. Experiments were repeated (n) times, where (n) is the number of images in the database; final classification accuracy of the system is the average of all n times.
- **70% Training and 30% Testing:** in this strategy, 70% of images (normal and abnormal) were used as training and the rest used for testing. For the mini-MIAS database, 146 of normal and 79 of abnormal images were used as training, while 63 of normal and 34 of abnormal images were used for testing. For the DDSM database, 179 of normal and 179 of abnormal images were used as training while remaining 77 images from each of normal and abnormal images were used as testing images. Images were divided into training and testing sets using random selection and the experiments were repeated 20 times. The reported results are the average result of 20 experiments

- **50% Training and 50% Testing:** One half of the data set used for training and the other half used as testing. For the mini-MIAS database, 105 of normal and 57 of abnormal images were used as training, while the remaining 104 normal and 56 abnormal images respectively were used for testing. For the DDSM database, 128 of normal and 128 of abnormal images were used as training set while remaining 128 images from normal and abnormal images were used as testing images. Again, the experiments were repeated 20 times by random selection of training images. Reported results are the averaged results of the 20 experiments.
- **10% Training and 90% Testing:** in this scenario, only 10% of images per class (normal and abnormal) were used as training and the rest as testing images. For the mini-MIAS database, 21 of normal and 11 of abnormal images were used for training, while 188 of normal and 102 of abnormal images were used for testing. For the DDSM database, 26 of normal and 26 of abnormal images were used as the training set while remaining 230 images from each of normal and abnormal images were used as testing images. Experiments were repeated 20 times using random sampling as before, and the average results are analysed for evaluation.

5.3.3 Classification Methods

To date, there is no optimum classifier that can fit every type of data. Therefore, in our work, we used one of the most commonly used classifier in the field of machine learning, which is the support vector machine (SVM). SVMs have shown to be a useful classifier, particularly when classifying data into one of two classes. We trained a binary SVM classifier to classify a given mammogram as normal or abnormal based on the above features.

5.4 Experimental results and discussion

In this section, we will present and discuss the results of the various experiments conducted to evaluate the performance of the proposed features. In each case, performance will be measured in terms of Sensitivity, Specificity, and Accuracy of the system. Details of the three measurements were given in Section 2.3.3.

5.4.1 Multi-Scale LBP

- **Single Sub-band:** In this case, only a single wavelet sub-band used to extract LBPH from it and represent whole mammogram image after mammogram images are decomposed into two levels using discrete wavelet transformations.

Experimentally we exclude the HH sub-band from the feature extraction process. Table 5.1 to Table 5.4 show the classification results of proposed technique applied on two used databases and four different strategies (Leave-One-Out strategy, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing,) respectively. Each table presents sensitivity, specificity, and accuracy of the system.

Table 5.1: Results of leave-one-out strategy based on LBPH features obtained from single wavelet sub-band

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL1	90.27	94.74	93.17	87.89	89.45	88.67
HL1	89.38	98.57	95.34	64.45	64.06	64.26
LH1	87.61	97.61	94.10	60.94	64.06	62.50
LL2	76.11	92.82	86.96	89.45	89.84	89.65
HL2	61.95	79.43	73.29	69.92	71.88	70.90
LH2	65.49	88.04	80.12	72.27	74.22	73.24

Table 5.2: Results of 70%/30% training/testing strategy based on LBPH features obtained from single wavelet sub-band

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL1	86.77	93.41	91.08	88.51	88.90	88.70
HL1	86.77	97.62	93.81	62.47	60.97	61.72
LH1	87.35	98.33	94.49	63.90	62.73	63.31
LL2	76.62	93.10	87.32	89.68	89.03	89.35
HL2	59.27	83.41	74.95	71.04	70.46	70.75
LH2	67.21	86.91	80.00	73.64	72.40	73.02

Table 5.3: Results of 50%/50% training/testing strategy based on LBPH features obtained from single wavelet sub-band

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL1	86.07	93.46	90.88	87.81	88.87	88.34
HL1	85.98	98.70	94.25	62.19	62.42	62.31
LH1	84.46	97.98	93.25	61.48	62.11	61.80
LL2	77.86	92.16	87.16	87.97	89.10	88.54
HL2	58.21	82.74	74.16	69.69	67.50	68.59
LH2	64.46	85.91	78.41	71.84	72.85	72.34

Table 5.4: Results of 10%/90% training/testing strategy based on LBPH features obtained from single wavelet sub-band

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL1	78.58	92.95	87.90	83.15	85.78	84.47
HL1	72.84	98.35	89.38	55.96	58.44	57.20
LH1	74.66	98.67	90.22	57.35	54.94	56.14
LL2	68.24	91.38	83.24	87.28	86.11	86.70
HL2	41.62	85.75	70.22	63.30	65.37	64.34
LH2	44.41	90.08	74.02	65.80	66.37	66.09

From the results obtained above we can notice that:

- Comparing different experimental scenarios for both databases shows that the performance of each of feature vector schemes follows a very similar pattern when comparing the various wavelet sub-band representations across all the 4 scenarios. The leave-one-out yields better classification accuracy than other 3 remaining scenarios, 70/30 training/testing scenario yields better classification accuracy than 50/50 training/testing and 10/90 training/testing scenarios, and 50/50 training/testing have better classification accuracy than 10/90 training/testing. These patterns are not surprising because the larger the gallery, the better chance for accurate matches.

- The best classification accuracy rates achieved for the different scenarios are as following:
Mini-MIAS:Leave-one-out:95.342%(HL1); 70/30 training/testing:94.485%(LH1); 50/50 training/testing:94.250%(LH1); 10/90 training/testing:90.224 %(HL1).
DDSM:Leave-one-out:89.648%(LL2); 70/30 training/testing: 89.351%(LL2); 50/50 training/testing: 88.535%(LL2); 10/90 training/testing: 86.696%(LL2).
- **Multi Sub-bands:** Here in this section we show the results of combining more than one wavelet sub-bands. The combination of sub-bands results from the fact that a single sub-band may not represent mammogram images in a good manner as gathered from the results shown in Table 5.1 to Table 5.4 in the previous section. Also, we conclude from the previous section that first level of wavelet decomposition has higher accuracy than the second level of the wavelet decomposition especially for Mini-MIAS database, therefore, in this case, we only combine the first level sub-bands. Table 5.5 to Table 5.8 shows results of different sub-bands combination with four different strategies applied on two databases.

Table 5.5: Results of leave-one-out strategy based on LBPH features obtained from multi wavelet sub- bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	94.69	99.04	97.52	87.50	89.06	88.28
LL,LH	96.46	99.04	98.14	88.67	90.63	89.65
LL,HL,LH	94.69	99.52	97.83	88.28	89.84	89.06
ALL	94.69	99.52	97.83	88.67	91.02	89.84

Table 5.6: Results of 70%/30% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	92.35	96.91	95.31	88.77	88.70	88.73
LL,LH	92.65	97.54	95.83	89.29	89.81	89.55
LL,HL,LH	92.65	98.41	96.39	89.16	89.87	89.51
ALL	94.12	99.37	97.53	88.57	90.33	89.45

Table 5.7: Results of 50%/50% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	89.38	96.39	93.94	86.84	88.87	87.85
LL,LH	90.00	97.02	94.56	87.46	90.12	88.79
LL,HL,LH	89.82	97.89	95.06	86.56	89.65	88.11
ALL	92.59	98.80	96.63	86.88	89.65	88.26

Table 5.8: Results of 10%/90% training/testing strategy based on LBPH features obtained from multi wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	81.23	94.92	90.10	83.72	86.00	84.86
LL,LH	81.32	94.97	90.17	85.33	86.17	85.75
LL,HL,LH	81.91	96.04	91.07	85.24	86.26	85.75
ALL	84.36	97.34	92.78	85.15	86.52	85.84

From results gathered in this section which shown in Table 5.5 to Table 5.8 and by comparing them with obtained results in Table 5.1 to Table 5.4 respectively, it is clear that results improved when LBP features of one or more non-LL sub-band(s) were combined with LBP features of LL-sub-band. This shows that using single sub-bands to represent mammogram images is not sufficient and each sub-band contain useful discriminating features and can play its role when more than one sub-bands are combined to represent mammogram images. This is due to the multi-resolution characteristics of wavelet transforms, which mean that significant texture features that hold discriminating information, appear at diferent wavelet frequency subbands with varying weights (i.e. wavelet coeffocients) By fusing more than subbands, we reduce the possibility of lossing discriminating information. The LL subbands, being an approximation representation of the original images, for example are outperformed by most of higher frequency subbands.

- **Original with sub-band(s):** Here in this section we propose another combination between LBPH features extracted from original mammogram image in the spatial domain and LBPH features extracted from one or more wavelet-sub-bands. The idea comes out from the fact that LBPH features extracted in the spatial domain had good classification results when they used alone see Table 5.9. On the other hand, we proved and obtained good classification accuracy when LBPH features are extracted from wavelet sub-band(s). Therefore, to obtain higher classification accuracy we combine two LBPHs. Table 5.10 to Table 5.13 shows results of combination with different sub-band(s) and different strategies.

Table 5.9: Results of different strategy based on LBPH features obtained from original mammogram images

	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	92.92	96.65	95.34	88.67	92.58	90.63
70%/30% training/ testing	91.07	95.69	94.08	90.07	91.43	90.75
50%/50% training/ testing	90.18	95.22	93.46	88.98	90.51	89.75
10%/90% training/testing	88.24	94.15	92.07	82.72	86.76	84.74

Table 5.10: Results of leave-one-out tests strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	97.35	99.04	98.45	90.63	92.58	91.60
O,HL	96.46	99.04	98.14	90.23	91.41	90.82
O,LH	95.58	99.04	97.83	91.02	92.19	91.60
O,LL,HL	97.35	99.04	98.45	90.23	92.58	91.41
O,LL,LH	97.35	99.04	98.45	90.23	93.36	91.80
O,LL,HL,LH	97.35	99.04	98.45	89.84	93.75	91.80
O,ALL	97.35	99.52	98.76	89.84	93.75	91.80

Table 5.11: Results of 70%/30% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	96.77	99.21	98.35	89.74	92.53	91.14
O,HL	95.74	99.13	97.94	89.29	91.17	90.23
O,LH	95.59	98.97	97.78	89.55	91.69	90.62
O,LL,HL	96.47	99.21	98.25	89.74	92.66	91.20
O,LL,LH	96.47	99.21	98.25	90.00	92.86	91.43
O,LL,HL,LH	96.47	99.21	98.25	90.07	92.53	91.30
O,ALL	96.32	99.21	98.20	90.33	92.47	91.40

Table 5.12: Results of 50%/50% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	95.36	98.80	97.59	89.92	90.66	90.29
O,HL	95.00	98.75	97.44	89.02	90.90	89.96
O,LH	95.09	98.75	97.47	89.10	91.21	90.16
O,LL,HL	95.45	98.94	97.72	89.73	90.63	90.18
O,LL,LH	95.27	98.99	97.69	89.77	90.78	90.27
O,LL,HL,LH	95.27	98.99	97.69	89.65	90.86	90.25

O,ALL	95.27	98.99	97.69	89.45	91.13	90.29
--------------	-------	-------	-------	-------	-------	-------

Table 5.13: Results of 10%/90% training/testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Wavelet Sub-bands	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	91.28	98.88	96.21	84.00	87.48	85.74
O,HL	91.13	98.83	96.12	82.83	87.02	84.92
O,LH	91.18	98.88	96.17	83.13	86.87	85.00
O,LL,HL	91.32	98.88	96.22	84.04	87.52	85.78
O,LL,LH	91.28	98.88	96.21	84.24	87.50	85.87
O,LL,HL,LH	91.28	98.91	96.22	83.04	87.20	85.12
O,ALL	91.47	98.99	96.35	84.35	87.65	86.00

Results obtained from Table 5.10 to Table 5.13 can be compared in two ways; first, if we compare with results shown in Table 5.9 which is the results of extracting LBPH from the original mammogram images only, we can notice a very high improvement of the classification accuracy results. Secondly, if we compared with results shown in Table 5.1 to Table 5.8 which are results of using single and multi sub-bands, again we notice that results are highly improved using the new combination strategy which is combining the extracted features from original mammogram image with extracted features from wavelet sub-band(s). These results can be explained partly by the explanation, we gave earlier, before the current group of tables, regarding the appearance of significant discriminating texture feature over multiple subbands. In fact, even fusing the original LBPH with that of a single wavelet subband seem to compensate for the loss of information when using the single subband.

5.4.2 GLCM

To test the performance of the GLCM features, as explained in (Section 5.2.2), extracted from the cropped ROI version of the images in the four different angles (0, 45, 90, and 135). The testing was conducted with the four different strategies of separating test and train sets. Table 5.14 to Table 5.17 show individual results for the 4 different angles, and the 2 databases. Except for the 10%/90% training/testing protocol, the obtained results show high accuracy especially for the leave-one-out strategy. In all case, the performance of the GLCM scheme is particularly high for the Mini-Mias images, which can be attributed to the differences between the images in the two databases. These results also confirm that using the ROI mammogram images give better classification accuracy when compared with the results of using whole images or even compared with our previously proposed method for removing unwanted image parts shown in the previous chapter.

Table 5.14: Results of leave-one-out tests strategy based on 14 GLCM features obtained from the original mammogram

Angle	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
R0	96.46	100.00	98.76	97.27	91.02	94.14
R45	94.69	100.00	98.14	88.67	85.55	87.11
R90	96.46	100.00	98.76	97.27	93.36	95.31
R135	96.46	100.00	98.76	96.48	92.58	94.53

Table 5.15: Results of 70%/30% training/testing strategy based on 14 GLCM features obtained from the original mammogram

Angle	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
R0	86.03	98.10	93.87	93.29	90.07	91.68
R45	87.79	97.54	94.12	95.92	80.07	87.99
R90	74.85	95.24	88.09	96.97	86.71	91.84
R135	85.29	98.73	94.02	95.20	88.36	91.78

Table 5.16: Results of 50%/50% training/testing strategy based on 14 GLCM features obtained from the original mammogram

Angle	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
R0	91.70	98.46	96.09	92.27	90.90	91.58
R45	92.41	98.03	96.06	93.01	76.48	84.75
R90	92.50	89.76	90.72	95.98	83.95	89.96
R135	92.05	99.04	96.59	93.75	84.69	89.22

Table 5.17: Results of 10%/90% training/ testing strategy based on 14 GLCM features obtained from the original mammogram

Angle	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
R0	82.26	71.30	75.16	82.50	69.39	75.95
R45	80.49	83.96	82.74	79.48	62.52	71.00
R90	71.81	73.22	72.72	82.37	59.09	70.73
R135	74.17	78.94	77.26	80.85	60.59	70.72

5.4.3 HOG

Table 5.18 shows the accuracy of the system when using the HOG features as texture representation of mammogram images. The results are given for the two different databases for the 4 different strategies of evaluation. Compared to the previous methods, the HOG-based scheme has a disappointingly low performance. This can be attributed to the very large feature vector size. Therefore it would of great interest to test whether this scheme can benefit or not from feature reduction. Details of feature reduction methods and experiments will be given in Section 5.6. Again there are significant differences between the performance of the scheme for the 2 different databases of mammograms.

Table 5.18: Results of using four strategies based on HOG features obtained from original mammogram

	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	80.53	98.09	91.93	72.66	77.34	75.00
70%/30% training/ testing	80.00	98.02	91.70	71.56	78.57	75.07
50%/50% training/ testing	77.95	98.17	91.09	69.49	80.00	74.75
10%/90% training/testing	56.52	98.59	83.79	60.13	77.85	68.99

5.5 Results Summary

We chose the best classification accuracy obtained by each individual feature type from the results shown in Table 5.8 to Table 5.18 to compare the different feature extraction methods proposed in previous sections. Figure 5.6 and Figure 5.7 show best classification results for Mini-MIAS and DDSM database respectively:

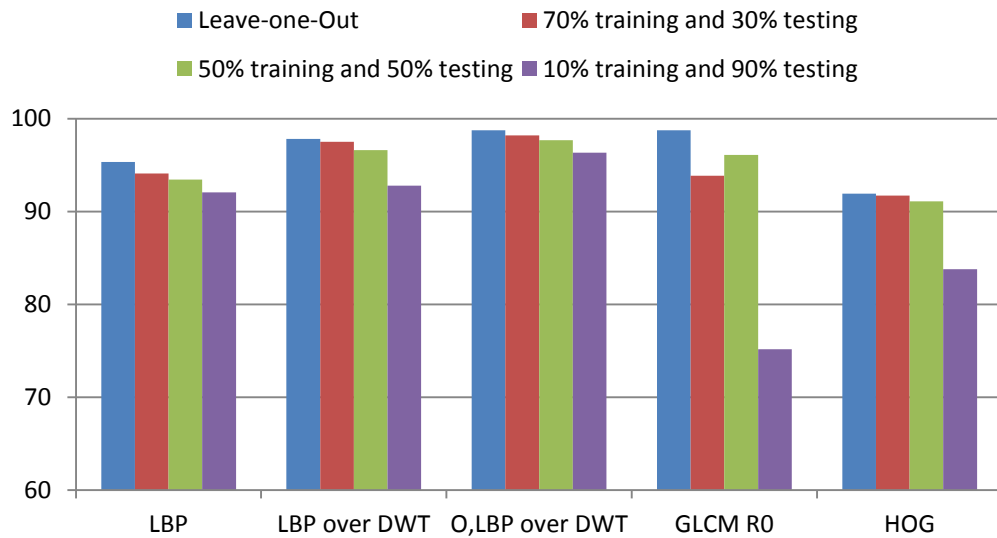


Figure 5.6 : Results summary of Mini-MIAS database accuracy

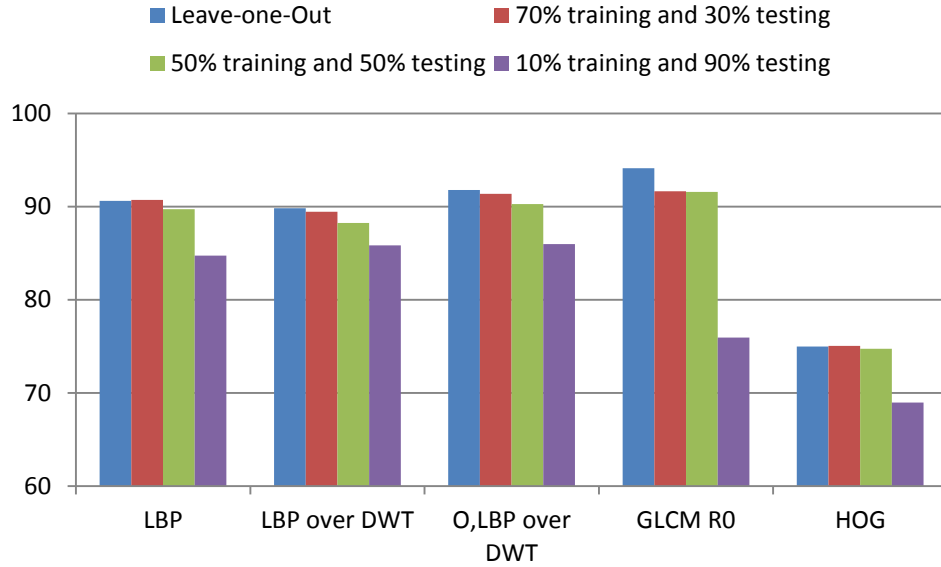


Figure 5.7 : Results summary of DDSM database accuracy

Results in Figure 5.6 and Figure 5.7 show that combining LBPH features from the original image with LBPH features extracted from wavelet sub-bands produced a higher classification accuracy than the respective individual LBPH features and HOG features. The reason for this is that LBP codes reflect the order relation between all the neighbouring pixels and the centre pixel in the local neighbourhoods rather than variations in pixel values in specific directions. Also, the LBP feature would be invariant under monotonic grey level changes, i.e. the individual greylevel values hold less significant information about the breast tissue models. As explained before, extracting LBP from wavelet sub-band(s) benefits from the multi-resolution nature of wavelet transforms. One another reason is that in the proposed scheme we combined LBPH features extracted from the spatial domain image directly and LBPH features extracted from different wavelet sub-bands so that the obtained feature vector will have a better representation of the texture at different resolution of the mammogram images. The achieved high accuracy of our system using LBPH is due to the sensitivity of the method to the mammogram images structure which contain mixture of complex fine lines and dense areas compared to other medical images type like ultrasound.

Although, not as good as LBP features, GLCM features also reflect/highlight variations in neighbouring mammogram pixel values in selected directions. This means that GLCM also encapsulate image textures which seem to have good discriminating representation of mammograms. Comparing GLCM with multi-scale LBP, GLCM

provides features in the directions 45, 135 which is not represented in the multi-scale LBP, and this made GLCM performed same or slightly better than Multi-scale LBP, particularly when a large number of training samples are used (Leave-One-Out strategy). This proved that the fourteen well known statistical features obtained from co-occurrence matrix have a very good representation of the correlation in the intensity of pixels (grey levels) that are next to each other appeared in breast tissue.

On the other hand, HOG features produced the lowest classification accuracy compared with all other used feature vectors. While in our case, we did not consider mass objects types and shapes which might be different from one abnormal mammogram to another and we took all abnormal cases as one type. Therefore, the results obtained using HOG descriptor was expected to be low because the HOG descriptor originally has been successfully used in object detection, for instance, in human face detection or recognition. Another reason could be the feature vector size that represents the mammograms, for the HOG descriptor has 8100 features which are relatively high. Also the information of the mammogram image is different from other types of images, so this affect the accuracy of the system.

Therefore, our next focus will be reducing feature vector size (see next section). One other point that needs mentioning is that classification results obtained from the Mini-MIAS database are higher than the classification results obtained from the DDSM database. One reason could be the structure of the database affect the results, for example, Mini-MIAS database contain a different type of abnormality that leads to classifying the cases to abnormal cases easier than DDSM database which contains only one type of abnormality.

5.6 Dimension Reduction

High dimensional feature vectors could contain redundant features that can have a negative effect on classification accuracy as well as computational efficiency. Redundant features can be described as those features that provide no more information than the currently selected features, and irrelevant features provide no useful information in any context (Tan, et al., 2006) (Hastie, et al., February 2009). Although the extracted feature vector is much smaller than the original image size, still the feature set can be seen as large and may contain redundancies. This general problem is

commonly addressed by process of feature selection, feature weighting, or dimension reduction (Oral & Sezg, 2013) (Suhas, et al., 2012).

There are many algorithms and techniques developed for dimensionality reduction such as Singular Value Decomposition (SVD), Independent Component Analysis (ICA), Principal Component Analysis (PCA), and Random Projection (RP) (Bingham & Mannila, 2001). While all of these methods have a similar goal, their approach to the problem is different. Here in this thesis, we use Principal Component Analysis (PCA) as a feature reduction method. Principle Component Analysis is a well-known and an effective dimension reduction technique used in many different classification problems.

The high dimensional feature vectors as proposed or investigated in the previous sections are in different sizes and could contain redundant features. For example, the number of features in the GLCM based feature representation is only 14 features; in multi-scale LBP, the number of features is 531 features for each original or wavelet sub-bands; while for HOG features the feature vector size is 8100 features. Table 5.19 summarises the number of features using different feature extraction techniques. The table also contains the size of using a different number of wavelet sub-bands in the case of LBP. We can see that all feature vectors, except GLCM features, are high dimensional and therefore, could benefit from dimension reduction.

Table 5.19: Feature vector size

Technique used	Wavelet sub-band(s)	Feature vector size
GLCM		14
LBP	One	531
	Two	1062
	Three	1593
	Four	2124
	Original and One	1062
	Original and Two	1593
	Original and Three	2124
	Original and Four	2655
HOG		8100

In this section, we investigate and propose the use of dimension reduction to reduce the size of feature vectors that represent mammogram images (or their ROIs).

Experimental results will demonstrate an enhancement in mammogram classification accuracy in most instances when dimension reduction is employed on the features proposed in previous sections. Dimension reduction using PCA, which is the method we use, will be explained in details here.

Principal components analysis (PCA) is one of the very popular techniques for dimensionality reduction. The main idea of PCA is to find a lower dimensional representation of a high dimensional data set that consist a large number of interrelated variables (features). Data variables are mostly correlated to each other unless the structure of the data is not simple or the dimension is small (Jolliffe, 2002). This characteristic means that discarding some of these variables/features might not affect characterizing the samples. PCA offers a linear transformation of the data from its original space into a de-correlated space. The newly offered space has covariance close to zero between each pair of dimensions. Computation of PCA is based on finding a matrix, in which its columns are the Eigenvectors of the data covariance. The Eigenvalue λ corresponding to the Eigenvector v of the data covariance C is equivalent to data variance in the direction of v . For this reason, the direction of k Eigenvectors corresponding to the biggest k eigenvalues cover as much data variance as possible by k orthogonal components (Mika, et al., 1999). The Eigenvectors that have small corresponding Eigenvalues could be discarded because the projected data on those components are close to zero Figure 5.8. The following explains the computation steps for data transformation using PCA:

Step 1: Suppose that the data is denoted by, $X = \{x_1, x_2, \dots, x_N\}$, $x_i \in \mathbb{R}^d$.

Step 2: Subtract the mean values from the data such that the mean value of the data shifted to a d dimension point 0.

Step 3: Compute the covariance matrix Σ of X .

Step 4: Compute the Eigenvectors of Σ , and arrange them as columns in transformation matrix T , such that as small as the index of the Eigenvector column; it has the highest correspondence Eigenvalues.

Step 5: Transform the data X into X' using the following formula:

$$X' = X T$$

Where, X' is the transformed data in the principal component space, note that the last columns of T could be discarded, because that it represent the directions with lowest sample variance.

PCA is used for dimension reduction of high dimensional data for producing a compressed version of the data through transforming them into some of the highest Eigenvectors of the data covariance. In other words, only the first k number of column in PCA components will be selected to represent high dimensional data. In this study, we used different k values to see how selecting different number of PCA components will affect the accuracy of the system. PCA serves as a general-purpose tool with applications ranging from information extraction over dimension reduction to data visualization. Here, we propose to use PCA to reduce the dimension of LBP and HOG features. GLCM features are already small in number.

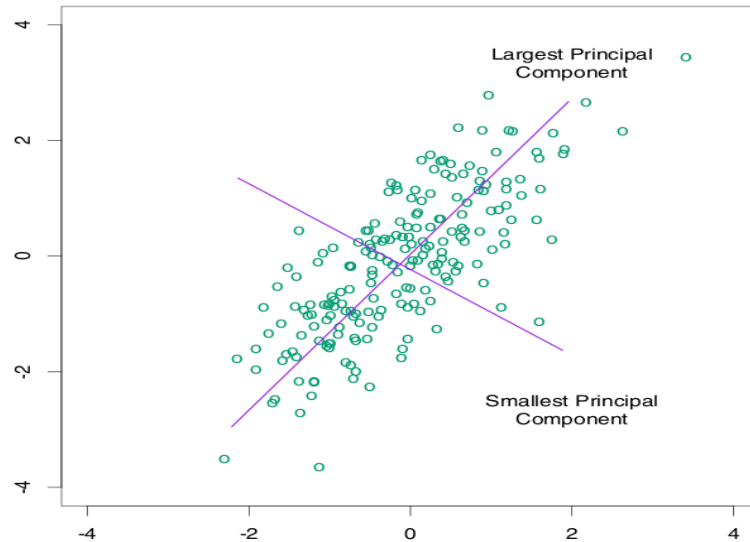


Figure 5.8: Largest and smallest Principle components

5.7 Dimension Reduction on Proposed Schemes

5.7.1 PCA over HOG

The HOG features, as extracted from the previous sections, include 8100 features. These features are probably too many and are redundant. Therefore, the PCA can be used to reduce the dimensionality of the feature vector and improve the classification accuracy. An additional benefit of dimension reduction is the reduction in the systems' computation cost, which could be useful in real-life situations when an automated system has to classify thousands of mammograms as quickly as possible.

Figure 5.9 illustrates the process of obtaining the low-dimensional HOG features by apply PCA on the set of HOG features obtained from a given mammogram or its ROI. We select the first k components that correspond the highest k eigenvalues. We call the new feature vectors PCA-HOG feature vectors.

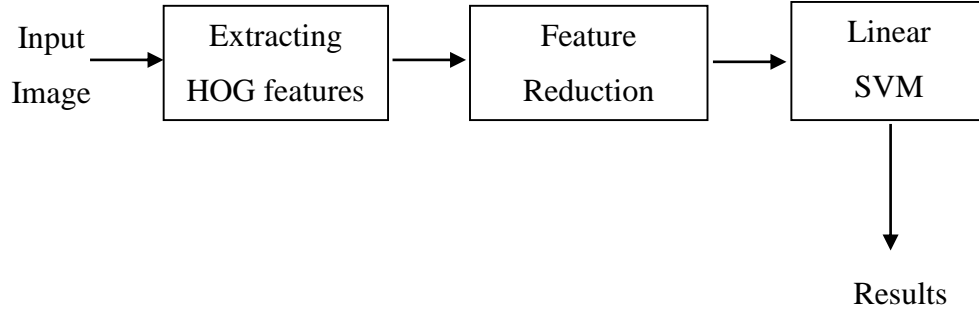


Figure 5.9: General frame work of PCA-HOG

5.7.2 PCA over LBP

We use PCA on LBP features in the same manner to the way PCA was used on HOG features to reduce the dimensionality of the feature space. The first k components that correspond to the largest k eigenvalues are selected as the reduced dimension features which we refer to as PCA-LBP.

5.7.3 Experimental Setup and Results

We used the same databases and the four different evaluation strategies (Leave-One-Out, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing) as before to evaluate our proposed PCA-HOG and PCA-LBP features. The binary SVM classifier was used for classification, and the results are reported in terms of sensitivity, specificity, and accuracy.

5.7.3.1 PCA over HOG

Table 5.20 to Table 5.23 show the classification sensitivity, specificity, and accuracy of the proposed PCA-HOG using four different evaluation strategies on the two different databases. In each case, the experiment is done using a different number of PCA components, and the results are compared with the full set of HOG feature (i.e. before applying PCA).

Table 5.20: Results of leave-one-out strategy based on applying PCA over HOG features with selecting various PCA components

	Mini-MIAS	DDSM
--	-----------	------

Number of PCA Components	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (8100)	80.53	98.09	91.93	72.66	77.34	75.00
1000	80.53	98.09	91.93	72.66	77.34	75.00
500	80.53	98.09	91.93	70.70	75.78	73.24
400	80.53	98.09	91.93	69.53	71.88	70.70
300	82.30	98.09	92.55	71.48	73.05	72.27
200	80.53	95.69	90.37	69.92	71.88	70.90
100	82.30	94.26	90.06	74.22	77.73	75.98
50	80.53	91.87	87.89	75.78	82.42	79.10
40	86.73	95.69	92.55	78.52	85.16	81.84
30	88.50	94.26	92.24	77.34	85.94	81.64
20	85.84	96.17	92.55	75.00	83.20	79.10
10	88.50	97.13	94.10	70.31	83.20	76.76

Table 5.21: Results of 70%/30% training/testing strategy based on applying PCA over HOG features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (8100)	80.00	98.02	91.70	71.56	78.57	75.07
1000	80.00	98.02	91.70	71.49	78.31	74.90
500	80.00	98.02	91.70	71.56	76.69	74.12
400	80.00	98.02	91.70	69.74	71.82	70.78
300	79.56	97.94	91.50	68.83	72.92	70.88
200	79.56	96.83	90.77	70.52	71.88	71.20
100	80.29	95.56	90.21	72.53	77.08	74.81
50	82.06	94.92	90.41	76.62	82.99	79.81
40	85.44	95.40	91.91	75.58	84.22	79.90
30	83.97	95.71	91.60	75.33	84.48	79.90
20	86.18	95.87	92.47	72.99	83.51	78.25
10	88.38	96.83	93.87	70.00	84.68	77.34

Table 5.22: Results of 50%/50% training/testing strategy based on applying PCA over HOG features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (8100)	77.95	98.17	91.09	69.49	80.00	74.75

1000	77.95	98.17	91.09	69.77	79.92	74.84
500	77.95	98.17	91.09	69.53	79.53	74.53
400	77.95	98.17	91.09	68.52	75.74	72.13
300	77.77	98.27	91.09	67.77	75.47	71.62
200	76.88	97.31	90.16	68.71	72.38	70.55
100	78.30	96.15	89.91	70.16	74.69	72.42
50	81.25	94.57	89.91	73.40	82.11	77.75
40	82.14	93.65	89.63	71.64	83.24	77.44
30	84.11	93.17	90.00	72.62	83.67	78.15
20	85.36	93.94	90.94	70.31	83.83	77.07
10	87.68	95.24	92.59	69.22	83.32	76.27

Table 5.23: Results of 10%/90% training/testing strategy based on applying PCA over HOG features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (8100)	56.52	98.59	83.79	60.13	77.85	68.99
1000	56.52	98.59	83.79	60.22	77.76	68.99
500	56.52	98.59	83.79	60.02	77.80	68.91
400	56.52	98.59	83.79	59.91	76.28	68.10
300	56.08	98.56	83.62	58.67	75.00	66.84
200	53.87	98.56	82.85	59.44	75.33	67.38
100	59.07	97.98	84.29	59.76	74.22	66.99
50	64.56	96.52	85.28	61.26	74.83	68.04
40	66.52	95.96	85.60	63.09	74.46	68.77
30	66.47	94.79	84.83	63.78	76.15	69.97
20	70.78	94.81	86.36	65.09	76.20	70.64
10	73.53	95.27	87.62	66.87	78.74	72.80

It is clear from the above results that using PCA to reduce the HOG feature space had a positive effect on mammogram classification accuracy. A closer look at the tables Table 5.20 to Table 5.23 show that the best results on the Mini-MIAS database are obtained when the feature size is reduced to only 10 PCA component. Meanwhile, on the DDSM database, PCA-HOG features with 10 to 40 PCA components achieved a higher accuracy than using all 8100 HOG features.

The increase in accuracy due to feature space reduction varied slightly across the two databases and the four different evaluation strategies. For example, on Mini-MIAS database, the increase is around 1 to 4 degrees, whilst it is 4 to 6 degrees on the DDSM database. A noticeable difference between the two databases is that whereas both sensitivity and specificity increased on the DDSM database if compared with using original feature vector, while only the sensitivity is increased on the Mini-MIAS database. One reason for this could be the balance between number of normal and abnormal cases in the DDSM database, while for Mini MIAS are not.

5.7.3.2 PCA over LBP

Table 5.24 to Table 5.27 show the classification sensitivity, specificity, and accuracy of the proposed PCA-LBP using four different evaluation strategies on the two different databases. In each case, the experiment is done using a different number of PCA components, and the results are compared with the full set of LBP feature (i.e. before applying PCA).

Table 5.24: Results of leave-one-out strategy based on applying PCA over LBPH features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (531)	92.92	96.65	95.34	88.67	92.58	90.63
500	90.27	99.52	96.27	71.09	94.92	83.01
400	90.27	99.52	96.27	71.09	94.92	83.01
300	90.27	99.52	96.27	71.09	94.92	83.01
200	90.27	99.52	96.27	71.09	94.92	83.01
100	90.27	99.52	96.27	70.70	94.92	82.81
50	90.27	99.52	96.27	70.70	94.92	82.81
40	90.27	99.52	96.27	70.70	94.92	82.81
30	90.27	99.52	96.27	69.92	94.53	82.23
20	89.38	99.52	95.96	71.09	94.14	82.62
10	89.38	99.52	95.96	73.83	92.97	83.40

Table 5.25: Results of 70%/30% training/testing strategy based on applying PCA over LBP features with selecting various PCA components

	Mini-MIAS	DDSM
--	-----------	------

Number of PCA Components	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (531)	91.07	95.69	94.08	90.07	91.43	90.75
500	88.82	99.52	95.77	67.92	93.44	80.68
400	88.82	99.52	95.77	67.92	93.44	80.68
300	88.82	99.52	95.77	67.92	93.44	80.68
200	88.82	99.52	95.77	68.05	93.44	80.75
100	88.82	99.52	95.77	67.86	93.44	80.65
50	88.68	99.52	95.72	67.79	93.64	80.71
40	88.82	99.52	95.77	68.05	93.44	80.75
30	88.82	99.52	95.77	67.79	93.57	80.68
20	88.68	99.44	95.67	67.66	92.92	80.29
10	88.24	99.44	95.52	68.51	91.95	80.23

Table 5.26: Results of 50%/50% training/testing strategy based on applying PCA over LBP features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
All (531)	90.18	95.22	93.46	88.98	90.51	89.75
500	88.48	99.52	95.66	64.14	93.28	78.71
400	88.48	99.52	95.66	64.14	93.28	78.71
300	88.48	99.52	95.66	64.18	93.28	78.73
200	88.48	99.52	95.66	64.14	93.32	78.73
100	88.48	99.52	95.66	64.14	93.24	78.69
50	88.48	99.52	95.66	64.02	93.28	78.65
40	88.48	99.52	95.66	63.98	93.24	78.61
30	88.48	99.47	95.63	63.67	93.20	78.44
20	88.39	99.52	95.63	63.98	92.27	78.13
10	88.13	99.33	95.41	63.83	92.31	78.07

Table 5.27: Results of 10%/90% training/testing strategy based on applying PCA over LBP features with selecting various PCA components

Number of PCA Components	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy

All (531)	88.24	94.15	92.07	82.72	86.76	84.74
500	68.43	100.0	88.90	41.00	92.78	66.89
400	68.38	100.0	88.88	41.00	92.78	66.89
300	68.43	100.0	88.90	41.00	92.78	66.89
200	68.38	100.0	88.88	40.98	92.78	66.88
100	68.38	100.0	88.88	40.98	92.83	66.90
50	68.24	100.0	88.83	41.11	92.76	66.94
40	68.19	100.0	88.81	40.85	92.78	66.82
30	68.09	100.0	88.78	40.94	92.72	66.83
20	68.09	100.0	88.78	41.96	92.33	67.14
10	67.50	100.0	88.57	41.67	92.20	66.94

Unlike the similar results PCA-HOG features produced on the two databases, PCA-LBP features produced contrasting results on the same databases. Therefore, we will discuss results for each database separately.

Looking at the results obtained from the Mini-MIAS database, we noticed that the overall classification accuracy is enhanced in the cases of leave-one-out, 70% training, and 50% training strategies. The enhancement is by one or two degrees when compared to using the original LBP feature set. The accuracy reduced by about 3 degrees for 10% training strategy. A closer inspection of the results shows that the reduction of LBP feature space by PCA led to a significant reduction in system sensitivity, but achieved a significant increase in system specificity, which leads to the overall increase in classification accuracy. Again the reason could be the imbalance between normal and abnormal cases in the database (i.e. the number of normal cases are about twice of the number of abnormal cases).

In terms of sensitivity and specificity, the effect of using PCA on LBP features is similar when using the DDSM database. However, the reduction in sensitivity was so high that it reduced the overall classification accuracy. Here the number of normal and abnormal cases are similar therefore the effect is also similar.

5.8 Conclusions

In this chapter, we explained and proposed the use of different texture features extraction methods such as the mutli-scale LBP, GLCM, and HOG to classify mammograms as normal or abnormal.

After evaluating the proposed techniques using two different databases and different evaluation strategies we can conclude the following:

1. Interestingly, the overall performance of the proposed approach that fused the LBP feature extracted from the original image and LBP features extracted from wavelet sub-bands is similar for the four evaluation strategies (i.e. leave-one-out, 70%/30% training/testing, 50%/50% training/testing, and 10% /90% training/testing).
2. Overall, the classification accuracies on the Mini-MIAS database are higher than the DDSM database. The following differences between the two databases could be contributing to the differences in their classification accuracy.
 - We used all 322 images of the Mini-MIAS database, whilst we used only 512 randomly chosen images for DDSM database. Ratios of images for the two different classes, i.e. normal and abnormal, are different in the two databases. There are 209 normal and 113 abnormal images with different types of abnormalities in the Mini-MIAS database. The number of normal cases is near twice the number of abnormal cases. This imbalance of the class cases affects overall the accuracy of the system by making the system less sensitive. For this reason, if we look at the results we can notice that the specificity of the system is much higher than the sensitivity. While for DDSM database the number of cases is equal i.e. there are 256 normal mammogram images and 256 abnormal images and abnormal cases are all in one type which they are the true mass type.
3. Mammogram classification accuracy can be improved by combining the traditional LBP features obtained from original mammograms with LBP features obtained from its wavelet transformed images/sub-bands.
4. Using statistical features from GLCM to represent mammogram images resulted in a better specificity of the system, which in turn had a positive effect on overall the accuracy rate of the system.
5. Initial results obtained from using HOG features are interesting when used for mammogram classification problems, therefore more analysis and investigations are needed to improve results obtained here in this chapter.

We have shown that multi-scale LBP is a good representation of mammogram images for the purpose of classifying mammograms as normal or abnormal cases. However, combining LBP features obtained at several scales and frequencies results in a very large feature set that is very likely to include redundancies. This could have a negative effect on the classification accuracy. Therefore, the work was continued in this chapter

on the use of feature selection/reduction to obtain the most discriminate features from a large feature set to represent mammogram images. Moreover, we proposed the use of PCA to reduce the HOG, and LBP feature space sizes/dimensions that represent mammogram images (or their ROIs). Experiments were carried out on two different databases using four different evaluation strategies.

Results demonstrated an overall enhancement in mammogram classification accuracy when PCA was used to reduce the HOG feature space. It was possible to achieve a similar if not better results with only 10 features based on PCA-HOG compared to using 8100 HOG features. Similarly, PCA helped reduce the LBP feature space by a significant amount without affecting the overall classification accuracy of the Mini-MIAS database. However, the outcome on the DDSM database was different in that the system sensitivity was greatly reduced even with a relatively large number of PCA-LBP features were used. This resulted in an overall lower classification accuracy with PCA-LBP features on the DDSM database.

The results in this chapter show that the different features we investigated represent mammograms or their ROIs in different ways; capturing different, and potentially complementary, information which results in different classification accuracies. This motivated the use of feature fusion to enhance the mammogram classification further. Feature and decision fusion have been successfully used in many classification problems. Hence, we devote the next chapter to investigate the use of fusion strategies as a mean of enhancing mammogram classification.

Chapter 6

Fusion for Mammogram Classification

Recent researches have shown that a combination of different feature representations could lead to an improvement in classification accuracy over the use of any single feature representation. The features that do not achieve a high accuracy themselves have shown to be valuable when fused with other features. Fusion techniques have been used successfully in a variety of classification problems. Examples include: biometric recognition (Ross & Govindarajan, 2005), (Chai, et al., 2006), (Sellahewa & Jassim, 2008), (Kong, et al., 2013), (Sabir, et al., 2014), (Chai, et al., 2015); multimedia (Atrey, et al., 2010); content-based image retrieval (Vieux, et al., 2012), (Al-Jubouri, 2015), image segmentation (Taher, 2014) (Taher, et al., 2014) and remote sensing (Yu, et al., 2006) .

Fusion can be performed at two levels: feature level fusion and decision-level fusion (Atrey, et al., 2010). In feature-level fusion, two or more extracted feature sets from the same data source are combined into a single larger size fused feature vector. In decision-level fusion, the result of several different individual classifiers is combined to arrive at a final classification. A typical approach to decision-level fusion is the majority voting scheme.

This chapter focuses on the fusion of the features we used earlier (i.e. HOG, GLCM, LBP) to enhance the mammogram classification accuracy. We will also explore decision-level fusion where we treat each feature on its own to arrive at an initial classification. The rest of this chapter is organized as follows. Section 6.1 explains the two fusion schemes in detail. The proposed feature and decision level fusion schemes are presented in Section 6.2. An analysis of experimental results obtained after feature and decision level is given in Section 6.3. A further analysis of the significance of these experimental results using a chi-square test is presented in Section 6.4. Finally, Section 6.5 offers a brief conclusion to the work presented in this chapter.

6.1 Fusion Techniques

6.1.1 Feature-Level Fusion

The aim of feature fusion is to introduce additional discriminative information for classification. Here, the word fusion means to concatenate two or more feature sets to create one ‘fused’ feature set that represents the data in question (e.g. mammogram images or their ROI). The idea stems from the observation that each type of feature(s) may represent a specific set of discriminatory information about the original data, although one or few on their own could be far better than others in being able to represent the data. However, if performed correctly, combining different types of features into one feature set could enhance the discriminative power of the ‘best’ individual feature type since different feature types often capture complementary information about the original data. Therefore, feature-level fusion is expected to provide better classification results. Figure 6.1 illustrates the general process of feature-level fusion.

However, when we think about fusion at this level we need to consider: (i) the feature set size of multiple feature types may be incompatible resulting in one feature type dominating over the other (e.g., number of extracted features using HOG transform is 8100 features while using GLCM is only fourteen features); (ii) the data type of each feature set could be different from the others, in which case we have to normalize the features; and (iii) concatenating two or more feature vectors may result in a feature vector with very large dimensionality leading to the curse of dimensionality problem. This could degrade the system’s classification accuracy. Furthermore, large feature vectors increase the storage cost and require more computation time to process them, which is not ideal for a high-throughput system. However, the problem of high dimensionality can be overcome by performing feature transformation or feature selection.

6.1.2 Decision-Level Fusion

In a decision-based fusion of multiple classifiers, individual contributing classifiers are assumed to be competitive (i.e. their decisions tend to agree more than disagree) rather than being complementary. For this reason, each component classifier contributes to the final decision of classifying an input test sample often through a majority weighted rule.

Majority voting is a popular technique to combine outputs of multiple classifiers. Assume that n classifiers are deployed, and that for each input sample, each classifier produces a unique decision regarding the classification of the sample. In our work, this unique decision could be normal or abnormal. In combining the decisions of the classifiers, the sample is assigned to the class (normal or abnormal) when at least k classifiers agree on the class, where

$$k = \begin{cases} \frac{n}{2} + 1 & \text{if } n \text{ is even} \\ \frac{n + 1}{2} & \text{if } n \text{ is odd} \end{cases} \quad (6.1)$$

For this kind of fusion, there are two different scenarios. In the first scenario, the outputs of different classifiers of the same feature set are fused to generate a final classification result. The aim of this scenario is to see how using different classifiers on the same feature set can affect the result. While the second scenario fuses classification results of the same classifier, but with different feature sets. This latter scenario highlights the effect of using different feature sets on the overall classification accuracy. Figure 6.2 illustrates a general framework for decision fusion.

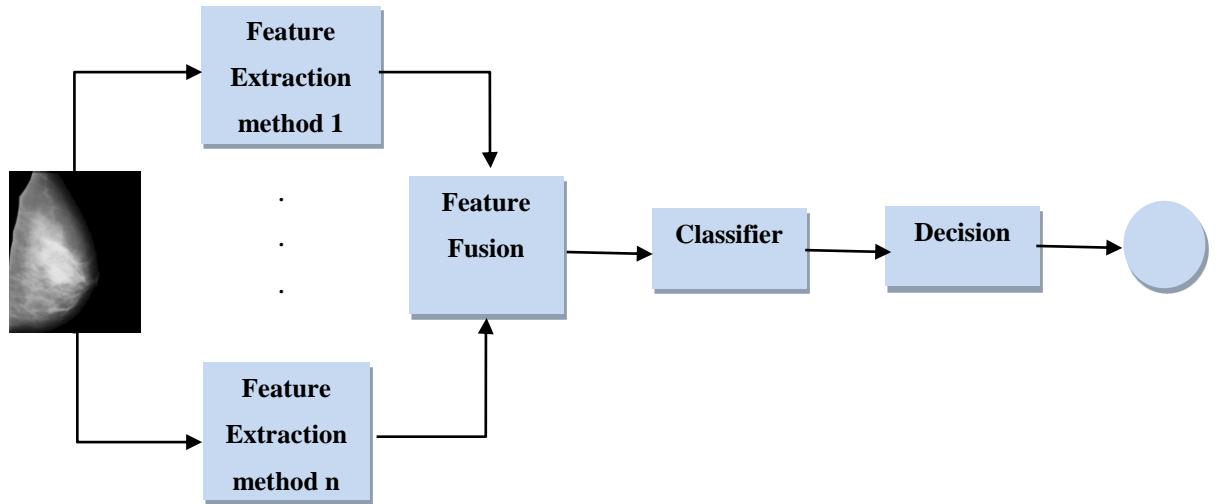


Figure 6.1: General frame work of feature level fusion strategy

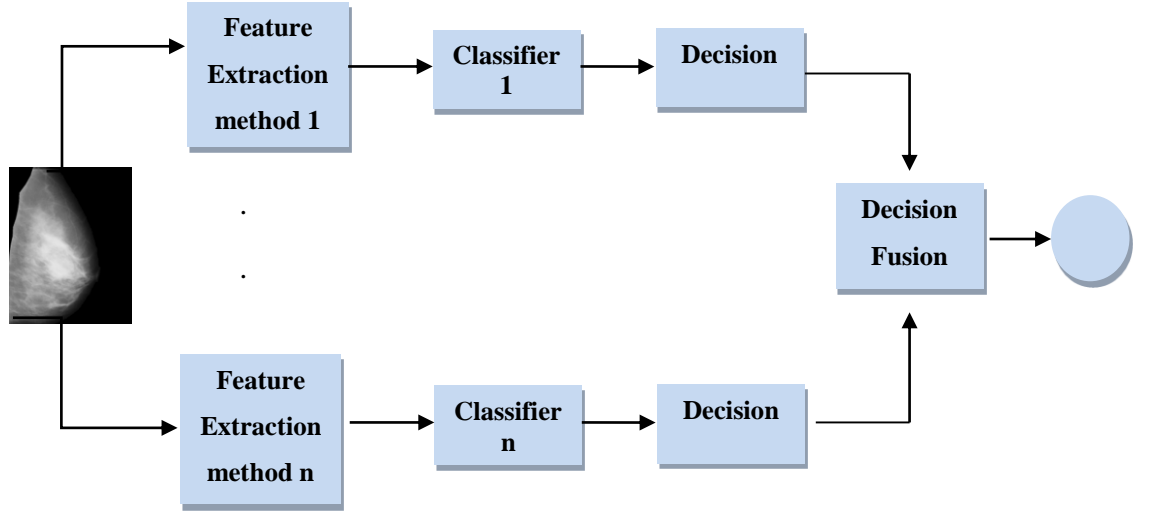


Figure 6.2: General frame work of the decision level fusion strategy

6.2 Feature and Decision Fusion for Mammogram Classification

Results in Chapter 5 showed that the different features we investigated represent mammograms or their ROIs in different ways; the features captured different, and potentially complementary, information which resulted in different classification accuracies. This motivated us to consider the use of features as well as decision fusion to enhance the mammogram classification further. This section explains our proposals for feature-level fusion and decision-level fusion to classify mammogram images as normal or abnormal images. Different fusion strategies are considered and evaluated using the same criteria and the databases use in Chapters 5.

6.2.1 Feature-Level Fusion for Mammogram Classification

The feature-level fusion is simply a concatenation of two or more of extracted feature vectors from the same mammogram image. Different ways of fusion and different feature types have been adopted to understand the effect of feature fusion compared with using single feature representation.

6.2.1.1 Fusion of HOG and LBPH features

In this scenario, we fuse HOG features with the LBPH features by concatenating them into a single feature representation. LBPH features are extracted from the original mammogram image as well as from the wavelet sub-bands (see Figure 6.3). Since the values of the two feature sets are in different ranges, we normalised LBPH features by dividing vector elements over the norm of the vectors to be real numbers between 0 and 1 before concatenating them with HOG features, which are between 0 and 1.

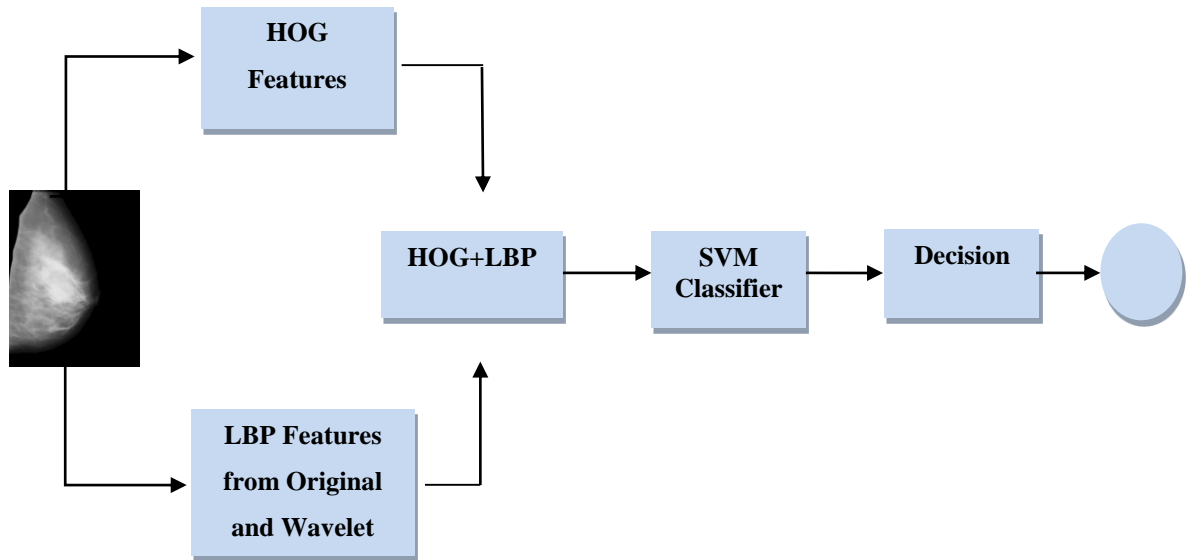


Figure 6.3 : Proposed HOG and LBPH features fusion diagram

Table 6.1 Table 6.1: Results of leave-one-out tests strategy based on fusing HOG and LBPH features to Table 6.4 below show the classification results of the two databases using four different experiment strategies. Different fusion strategies and the number of features after fusion are given in the tables.

Table 6.1: Results of leave-one-out tests strategy based on fusing HOG and LBPH features

Feature	Spatial domain / Wavelet Sub-band(s)	Number of Features	Mini-MIAS			DDSM		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
HOG	Spatial domain	8100	80.53	98.09	91.93	72.66	77.34	75.00
HOG+LBPH	O	8631	83.19	98.57	93.17	73.83	77.34	75.59
	LL	8631	80.53	98.09	91.93	75.00	79.69	77.34
	HL	8631	80.53	98.09	91.93	72.66	77.34	75.00
	LH	8631	80.53	98.09	91.93	72.66	77.34	75.00
	LL, HL	9162	80.53	98.09	91.93	73.44	77.34	75.39
	LL, LH	9162	80.53	98.09	91.93	73.44	77.34	75.39
	LL, HL, LH	9693	80.53	98.09	91.93	73.83	77.34	75.59
	ALL Sub-bands	10224	80.53	98.09	91.93	73.05	77.34	75.20
	O, LL	9162	84.07	98.57	93.48	75.00	80.47	77.73
	O, HL	9162	83.19	98.57	93.17	73.83	77.34	75.59
	O, LH	9162	83.19	98.57	93.17	73.83	77.34	75.59
	O, LL, HL	9693	84.07	98.57	93.48	73.83	78.52	76.17
	O, LL, LH	9693	84.07	98.57	93.48	73.83	78.52	76.17
	O, LL, HL, LH	10224	84.07	98.57	93.48	73.83	78.13	75.98
	O, ALL Sub-bands	10755	83.19	98.57	93.17	73.83	77.34	75.59

Table 6.2 : Results of 70%/30% training/ testing strategy based on fusing HOG and LBPH features

Feature	Spatial domain / Wavelet Sub-band(s)	Number of Features	Mini-MIAS			DDSM		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
HOG	Spatial domain	8100	80.00	98.02	91.70	71.56	78.57	75.07
HOG+LBP	O	8631	81.77	98.65	92.73	72.40	78.77	75.58
	LL	8631	80.74	97.94	91.91	73.90	80.39	77.14
	HL	8631	80.44	97.94	91.80	71.49	78.31	74.90
	LH	8631	80.44	98.02	91.86	71.49	78.31	74.90
	LL,HL	9162	80.74	97.94	91.91	72.27	78.64	75.45
	LL,LH	9162	80.74	97.94	91.91	72.21	78.57	75.39
	LL,HL,LH	9693	80.74	97.94	91.91	72.01	78.44	75.23
	ALL Sub-bands	10224	80.74	97.94	91.91	71.88	78.44	75.16
	O, LL	9162	82.06	98.57	92.78	74.55	80.65	77.60
	O, HL	9162	81.77	98.57	92.68	72.47	78.83	75.65
	O, LH	9162	81.77	98.57	92.68	72.40	78.77	75.58
	O, LL,HL	9693	81.91	98.65	92.78	73.25	79.35	76.30
	O, LL,LH	9693	81.91	98.65	92.78	73.25	79.42	76.33
	O,LL,HL,LH	10224	81.77	98.65	92.73	72.73	79.16	75.94
	O,ALL Sub-bands	10755	81.77	98.65	92.73	72.60	79.09	75.84

Table 6.3 : Results of 50%/50% training/testing strategy based on fusing HOG and LBPH features

Feature	Spatial domain / Wavelet Sub-band(s)	Number of Features	Mini-MIAS			DDSM		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
HOG	Spatial domain	8100	77.95	98.17	91.09	69.49	80.00	74.75
HOG+LBP	O	8631	79.82	98.37	91.88	70.31	80.27	75.29
	LL	8631	78.75	98.13	91.34	71.25	81.17	76.21
	HL	8631	77.95	98.17	91.09	69.73	79.81	74.77
	LH	8631	77.95	98.17	91.09	69.73	79.81	74.77
	LL,HL	9162	78.13	98.17	91.16	70.20	80.12	75.16
	LL,LH	9162	78.13	98.17	91.16	70.12	80.20	75.16
	LL,HL,LH	9693	78.04	98.17	91.13	69.92	80.00	74.96
	ALL Sub-bands	10224	77.95	98.17	91.09	69.81	79.96	74.88
	O, LL	9162	80.27	98.37	92.03	71.99	81.88	76.93
	O, HL	9162	79.82	98.37	91.88	70.27	80.43	75.35
	O, LH	9162	79.82	98.37	91.88	70.31	80.31	75.31
	O, LL,HL	9693	80.09	98.37	91.97	70.70	80.74	75.72
	O, LL,LH	9693	80.09	98.37	91.97	70.66	80.74	75.70

Feature	Spatial domain / Wavelet Sub-band(s)	Number of Features	Mini-MIAS			DDSM		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
	O,LL,HL,LH	10224	79.91	98.37	91.91	70.63	80.51	75.57
	O,ALL Sub-bands	10755	79.82	98.37	91.88	70.51	80.43	75.47

Table 6.4 : Results of 10%/90% training/testing strategy based on fusing HOG and LBPH features

Feature	Spatial domain / Wavelet Sub-band(s)	Number of Features	Mini-MIAS			DDSM		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
HOG	Spatial domain	8100	56.52	98.59	83.79	60.13	77.89	68.99
HOG+LBP	O	8631	58.58	98.70	84.59	60.44	78.04	69.24
	LL	8631	57.35	98.56	84.07	61.02	78.17	69.60
	HL	8631	56.72	98.56	83.85	60.22	77.74	68.98
	LH	8631	56.77	98.56	83.86	60.22	77.74	68.98
	LL,HL	9162	56.81	98.54	83.86	60.37	77.87	69.12
	LL,LH	9162	56.86	98.54	83.88	60.37	77.87	69.12
	LL,HL,LH	9693	56.86	98.54	83.88	60.26	77.87	69.07
	ALL Sub-bands	10224	56.86	98.54	83.88	60.24	77.83	69.03
	O, LL	9162	57.35	98.56	84.07	61.26	78.52	69.89
	O, HL	9162	59.61	98.62	84.90	60.39	78.02	69.21
	O, LH	9162	58.68	98.72	84.64	60.41	78.02	69.22
	O, LL,HL	9693	58.73	98.72	84.66	60.65	78.17	69.41
	O, LL,LH	9693	58.82	98.72	84.69	60.65	78.17	69.41
	O,LL,HL,LH	10224	58.87	98.72	84.71	60.59	78.13	69.36
	O,ALL Sub-bands	10755	58.78	98.72	84.67	60.52	78.09	69.30

Combining HOG and LBPH features resulted in a small increase when compared with the accuracy rates achieved with HOG features alone (results in Chapter 5, Section 5.4.3). However, the fusion led to a significant decrease in accuracy when compared with the accuracy achieved by LBPH features alone. This could be a result of the large size of HOG features (8100 features) overpowering the discriminability of LBPH features (531 features). We propose to use PCA to reduce the feature dimensions to address the size imbalance between HOG and LBPH features.

6.2.1.2 Fusion of PCA-HOG and PCA-LBP

Fusing HOG and LBP features results in a very large feature vector. However, Chapter 5 showed the size of HOG and LBP features could be reduced to a very small feature

vector by applying PCA for each feature type. Here we propose to fuse the PCA-HOG and PCA-LBP features by concatenating them to produce a single feature vector see Figure 6.4. For each feature, we selected the same number of PCA components. Table 6.5 and Table 6.6 show the classification results after fusing the two extracted features with different feature sizes. The tables show results for Leave one out and 70% training 30% testing strategies. Results of the other two strategies can be found in appendix D.

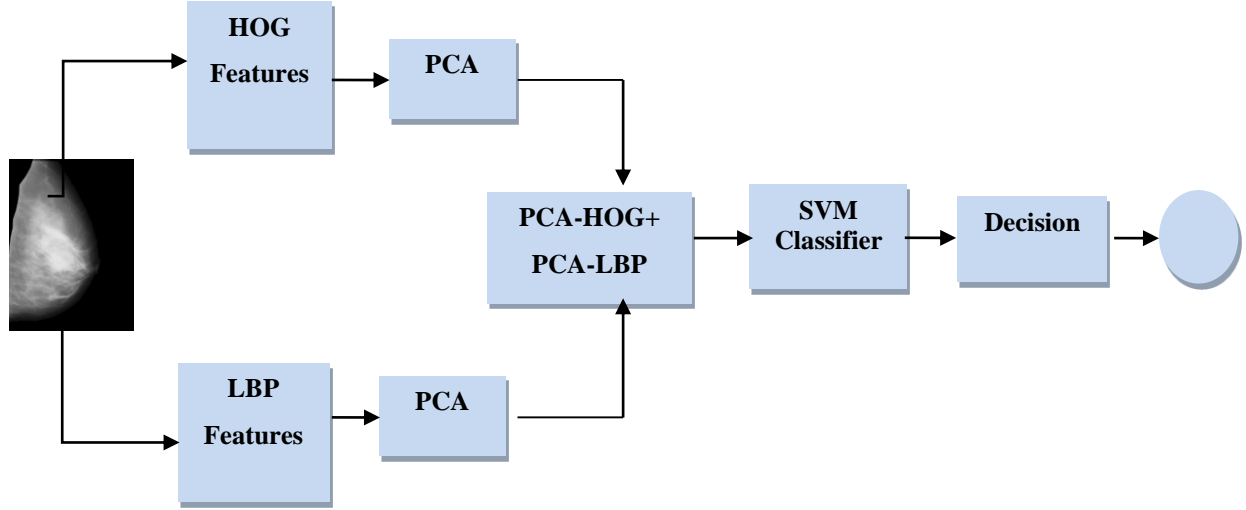


Figure 6.4 : Proposed PCA-HOG and PCA-LBPH features fusion diagram

Table 6.5 : Results of leave-one-out tests strategy based on fusing PCA-HOG and PCA-LBPH features

Selected Components	Final Feature size	Mini-MIAS			DDSM		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
500	1000	83.19	98.57	93.17	72.27	76.17	74.22
400	800	83.19	98.57	93.17	69.53	71.88	70.70
300	600	82.30	98.57	92.86	75.00	74.61	74.81
200	400	82.30	95.69	90.99	71.48	75.39	73.44
100	200	85.84	94.74	91.62	80.86	83.98	82.42
50	100	87.61	94.74	92.24	82.03	85.55	83.79
40	80	88.50	96.65	93.79	81.25	88.28	84.77
30	60	89.38	96.17	93.79	82.42	89.45	85.94
20	40	92.04	97.61	95.65	81.64	90.23	85.94
10	20	92.04	97.61	95.65	80.08	90.63	85.35

Table 6.6 : Results of 70%/30% training/testing strategy based on fusing PCA-HOG and PCA-LBPH features

Selected Components	Final Feature size	Mini-MIAS			DDSM		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
500	1000	81.77	98.57	92.68	72.47	77.47	74.97
400	800	81.77	98.57	92.68	70.20	72.47	71.33
300	600	82.21	98.49	92.78	69.42	73.25	71.33
200	400	80.88	97.22	91.50	71.30	71.36	71.33
100	200	82.35	96.43	91.50	75.00	78.70	76.85
50	100	85.59	95.64	92.11	80.13	85.84	82.99
40	80	87.94	96.43	93.45	78.57	87.08	82.83
30	60	86.47	96.91	93.25	80.84	87.47	84.16
20	40	89.85	97.70	94.95	80.13	89.22	84.68
10	20	90.44	98.41	95.62	76.69	89.87	83.28

The proposed feature fusion after dimension reduction led to a marginal improvement in classification accuracy when compared with the results of high-dimensional features stated in the previous set of experiments. Furthermore, the results achieved with fused features are slightly better than the results of individual features (after dimension reduction) as shown in the previous chapter. Nevertheless, the LBPH features taken from multiple wavelet sub-bands produced the best results. This led us to investigate fusion at decision level analogous to multiple experts examining a mammogram to arrive at a diagnosis.

6.2.2 Proposed Decision-Level Fusion methods

To overcome the problems of fusing large feature sets of different types, we propose to use decision-level fusion for mammogram classification. The proposed technique is as follows: for a given mammogram, we first find the classification result using different feature extraction methods. The classification is performed using an SVM classifier as explained before. We then use majority voting to arrive at the final decision to classify the mammogram as normal or abnormal. In all cases, we use three different feature sets to classify a given mammogram. Thus, for a given normal mammogram, if the classification result of two or more feature set is normal, the final decision is normal. Otherwise, the mammogram image will be classified as an abnormal case. On the other hand, for a given abnormal mammogram, if the classification result of two or more feature sets is abnormal, the final decision is then abnormal. Otherwise, the mammogram image will be classified as a normal case.

Table 6.7 and Table 6.8 show the classification results of fusing classification results of different feature sets for leave one out and 70%/30% strategy, respectively (see appendix D for other strategies). Experiments were conducted on the two mammogram databases using the four strategies of separating training and testing sets that we applied in previous sections. Table 6.7 and Table 6.8 shows the classification results of seven different decision fusing scenarios. For simplicity, we used $f1$ to $f7$ to represent different fusions as follows:

$f1$: fusion of LBP, HOG, and GLCM.

$f2$: fusion of LBP, HOG, and PCA-LBP+PCA-HOG

$f3$: fusion of LBP over O+LBP over W_ALL, PCA-HOG, and GLCM

$f4$: fusion of LBP over O+LBP over W_ALL, PCA-LBP+PCA-HOG, and GLCM

$f5$: fusion of LBP over O+LBP over W_ALL, PCA-HOG, and PCA-LBP+PCA-HOG

$f6$: fusion of PCA-HOG, PCA-LBP, and PCA-LBP+PCA-HOG

$f7$: fusion of PCA-HOG, PCA-LBP, and GLCM

Table 6.7 : Results of leave-one-out tests strategy based on decision level fusion of using a different kind of features

Fusions	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
$f1$	96.46	99.52	98.45	94.14	94.92	94.53
$f2$	92.04	98.57	96.27	83.59	91.41	87.50
$f3$	98.23	99.52	99.07	94.53	96.48	95.51
$f4$	97.35	99.52	98.76	94.14	96.48	95.31
$f5$	92.92	98.09	96.27	80.08	91.41	85.74
$f6$	92.04	98.09	95.96	78.13	90.63	84.38
$f7$	96.46	99.52	98.45	87.89	95.31	91.60

Table 6.8 : Results of 70%/30% training/testing strategy based on decision level fusion of using a different kind of features

Fusions	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
$f1$	93.97	99.37	97.47	94.48	97.08	95.78
$f2$	91.18	98.81	96.13	82.40	91.30	86.85

	Mini-MIAS			DDSM		
Fusions	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f3</i>	95.15	99.29	97.84	94.55	97.53	96.04
<i>f4</i>	95.59	99.29	97.99	95.00	97.40	96.20
<i>f5</i>	92.65	98.73	96.60	76.30	90.33	83.31
<i>f6</i>	91.62	98.89	96.34	75.65	90.33	82.99
<i>f7</i>	92.79	99.52	97.17	85.71	97.27	91.49

Gathered from the results above we can say that most of the decision fusion schemes work well and obtain high classification accuracy but *f3* and *f4* have higher classification results compared with other fusing schemes or using single feature sets. Moreover, results proved and showed the effects of our proposed combined features (LBP over original image + LBP over wavelet sub-bands) as well as our proposed feature reduction schemes applied on HOG and LBP. Obtained conclusion is true for both used databases and all four used strategies of separating data to training and testing sets.

6.3 Statistical evaluation of Experiment Results

We used statistical measurements test to analyse the significance of mammogram classification accuracy improvements obtained by the proposed methods in this and the previous chapter. First, we used the chi-square (x^2) test to evaluate the classification results. Then we analysed the score value that shows how far individual mammogram cases are from the hyper-plane in the SVM classifier. Details of the measurements are described below, which is followed by our analysis.

6.3.1 Significance between the observed and expected model

The chi-square (x^2) test is suitable for categorical data and it has been used to evaluate the significance of differences between the observed and the expected model (see equation 6.2) based on the contingency table (Field, 2006) (Al-Jubouri, 2015).

$$x^2 = \sum \frac{(\text{observed}_{ij} - \text{Model}_{ij})^2}{\text{Model}_{ij}} \quad (6.2)$$

$$Model_{ij} = \frac{Total_{rowi} \times Total_{columnj}}{n} \quad (6.3)$$

Where n is the total number of observations.

We used the chi-square (χ^2) test between two different approaches to mammogram classification at a time to decide if the improvement of the mammogram classification accuracy between them is significant or not.

Table 6.9 shows the general example of two mammogram classification systems obtained using different feature extraction methods (Method1 and Method2) with two different categories.

Table 6.9: Contingency table shows the frequencies of 1s and 0s for Method1 and Method2

	Category ₁ (1s)	Category ₂ (0s)	Total _{row}
Method1	M ₁	N ₁	T ₁
Method2	M ₂	N ₂	T ₂
Total _{column}	T _{M1+M2}	T _{N1+N2}	T

M1 and M2 represent the number (frequencies) of mammograms that are correctly classified using Method1 and Method2 respectively. While N1 and N2 represent the number (frequencies) of mammograms that are incorrectly classified using Method1 and Method2 respectively. T1 and T2 represent the total number of correct and incorrectly classified mammograms using Method1 and Method2 respectively.

In our case, we use the chi-square (χ^2) test to determine the significance between selected proposed schemes with other basic methods or with each other at (p-value < 0.05) significance level and degree of freedom equal to 1.

Table 6.10 presents the significant differences between mammogram classification results obtained from using different sets of features using the Mini-MIAS database with Leave- one-out strategy and SVM classifier. Significant means the two feature extraction schemes been compared give different recognition accuracies, and the difference between them is large (i.e. the mammogram classification accuracy improved by one of them). Results are not significant means the classification accuracy between the two features are very close to each other (i.e. there is no difference or little difference between accuracies obtained).

From the Table 6.10, we conclude that mammogram classification result is significant at $P < 0.05$ between some of the proposed methods, but not all. For example, the results between the following methods are significantly different:

- $f4$ and (Fusion of PCA-HOG and PCA-LBP), or
- (LBP over O+LBP over W_ALL) and LBP, or
- $f3$ and HOG, or
- $f4$ and LBP, or
- $f7$ and LBP, or
- $f7$ and HOG

This demonstrates that the proposed use of decision fusion, in particular, outperforms the existing single feature based approaches. However, the results between the following methods are not significant:

- (PCA-HOG) and HOG, or
- (Fusion of PCA-HOG and PCA-LBP) and LBP, or
- (Fusion of PCA-HOG and PCA-LBP) and HOG, or
- $f3$ and GLCM.

We then conducted the same analysis on the results obtained for the DDSM database. Table 6.11 below present the results these chi-square (χ^2) tests.

Table 6.10: Chi-square (χ^2) test for mammogram image classification on Mini-MIAS database

		PCA-HOG	Fusion of PCA-HOG and PCA-LBP	$f3$	$f4$	$f7$	LBP	HOG	GLCM
LBP over O+LBP over W_ALL	χ^2	10.1449	5.7153	0.1444	0	0.1127	6.562	16.9216	0
	p -value	0.001447	0.016818	0.703919	1	0.737107	0.010418	0.000039	1
	Result	Significant	Significant	Not significant	Not significant	Not significant	Significant	Significant	Not significant
PCA-HOG	χ^2	-	0.7985	12.0479	10.1449	8.4828	0.4968	1.1707	10.1449
	p -value	-	0.371545	0.000518	0.001447	0.003585	0.480902	0.279256	0.001447
	Result	-	Not significant	Significant	Significant	Significant	Not significant	Not significant	Significant
Fusion of PCA- HOG and PCA- LBP	χ^2	-	-	7.3106	5.7153	4.3928	0.0361	3.8384	5.7153
	p -value	-	-	0.006855	0.016818	0.036092	0.849291	0.050091	0.016818
	Result	-	-	Significant	Significant	Significant	Not significant	Not significant	Significant
$f3$	χ^2	-	-	-	0.14444	0.5063	8.23	19.1015	0.1444
	p -value	-	-	-	0.703919	0.47675	0.00412	0.000012	0.703919
	Result	-	-	-	Not significant	Not significant	Significant	Significant	Not significant
$f4$	χ^2	-	-	-	-	0.1127	6.562	16.9216	0
	p -value	-	-	-	-	0.737107	0.010418	0.000039	1
	Result	-	-	-	-	Not significant	Significant	Significant	Not significant
$f7$	χ^2	-	-	-	-	-	5.1603	14.9452	0.1127
	p -value	-	-	-	-	-	0.023109	0.000111	0.737107
	Result	-	-	-	-	-	Significant	Significant	Not significant

Table 6.11: Chi-square (χ^2) test for mammogram image classification on DDSM database

		PCA-HOG	Fusion of PCA-HOG and PCA-LBP	$f3$	$f4$	$f7$	LBP	HOG	GLCM
LBP over O+LBP over W_ALL	χ^2	43.6963	10.5084	5.9303	3.0055	0.0128	0.4385	52.1663	2.1513
	p -value	0	0.001188	0.014883	0.082985	0.909818	0.507826	0	0
	Result	Significant	Significant	Significant	Not significant	Not significant	Not significant	Significant	Not significant
PCA-HOG	χ^2	-	12.3119	75.3504	65.7758	102.0012	36.0678	0.4322	62.2136
	p -value	-	0.00045	0	0	0	0	0.51092	0
	Result	-	Significant	Significant	Significant	Significant	Significant	Not significant	Significant
Fusion of PCA- HOG and PCA- LBP	χ^2	-	-	30.5119	23.8451	9.8082	6.7359	17.2589	21.4892
	p -value	-	-	0	0.000001	0.001737	0.009449	0.000033	0.000004
	Result	-	-	Significant	Significant	Significant	Significant	Significant	Significant
$f3$	χ^2	-	-	-	0.5159	6.4781	9.4586	85.6421	0.975
	p -value	-	-	-	0.4726	0.010921	0.002102	0	0.323438
	Result	-	-	-	Not significant	Significant	Significant	Significant	Not significant
$f4$	χ^2	-	-	-	-	3.4051	3.5665	75.6233	0.0731
	p -value	-	-	-	-	0.064995	0.058956	0	0.786867
	Result	-	-	-	-	Not significant	Not significant	Significant	Not significant
$f7$	χ^2	-	-	-	-	-	0.3015	50.7216	2.4928
	p -value	-	-	-	-	-	0.582931	0	0.11437
	Result	-	-	-	-	-	Not significant	Significant	Not significant

From Table 6.11 we conclude that mammogram classification results between some of the proposed schemes are significant at $P < 0.05$, but not all. For example, the difference is significant when we calculate the significance value between the following two feature sets:

- (LBP over O+LBP over W_ALL) and (PCA-HOG), or
- (LBP over O+LBP over W_ALL) and (Fusion of PCA-HOG and PCA-LBP), or
- (LBP over O+LBP over W_ALL) and $f3$, or
- (PCA-HOG) and (Fusion of PCA-HOG and PCA-LBP), or
- (PCA-HOG) and $f3$, or
- (PCA-HOG) and $f4$, or
- (PCA-HOG) and $f7$, or
- (Fusion of PCA-HOG and PCA-LBP) and $f3$, or
- (Fusion of PCA-HOG and PCA-LBP) and $f7$, or
- $f3$ and LBP.

On the other hand, the results between the following used feature sets are not significantly different:

- (PCA-HOG) and HOG, or
- (LBP over O+LBP over W_ALL) and LBP, or
- $f7$ and LBP.

As a result of the above evaluation of the significances of classification accuracy, we can conclude that only some of the proposed schemes led to performance increase that is significant. Nevertheless, considering the potential ramifications of misclassification or misdiagnosis, even minor improvements could be important.

6.3.2 SVM Score Analysis

To understand our results further, we analysed the ‘score’ values of SVM classifier. In the SVM classifier, the score value shows if the individual mammogram case is classified as normal or abnormal by calculating the signed distance between observation x and the decision boundary (Hyperplane) which ranges from $-\infty$ to $+\infty$. A positive score for a class indicates that x is within that class, a negative score indicates otherwise. From the score

values, we calculated mean (μ), variance (σ^2), and standard deviation (σ) of the two classes to see how mammograms of the two classes are near/far from each other.

Table 6.12 below shows the results obtained from a randomly selected experiment on the Mini-MIAS database, with our proposed fusion strategy that combines PCA-LBP with PCA-HOG.

Table 6.12: SVM score analysis for Mini-MIAS database

	Leave One Out		70%/30% Train/ Test		50%/50% Train/Test		10%/ 90%Train/ Test	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
μ	1.974	3.574	1.681	2.584	1.931	3.203	0.960	0.936
σ^2	0.772	2.950	0.614	2.349	0.955	3.414	0.573	1.982
σ	0.597	8.702	0.377	5.516	0.912	11.655	0.328	3.929

The result in Table 6.12 shown above indicates that normal mammogram cases are relatively close to the hyper-plane than the abnormal cases. However, the normal cases are clustered in close proximity to each other and are separated from the abnormal cases by the hyper-plane. Whilst some abnormal cases are far away from the hyper-plane, they are scattered away from each other and are likely to be misclassified as a normal case. This could be due to the fact that abnormal mammogram cases include a wide range of abnormalities inside the database (e.g. calcifications; well-defined, spiculated or ill-defined masses; architectural distortion or asymmetry). Also, the severity of each abnormality is different (benign or malignant). However, we have included all different abnormalities as one class (Abnormal) to be compared against the Normal class. Therefore, what should be done is to separate the abnormal class into several different classes and train a multi-class SVM or set of binary SVMs to tackle a multi-class problem.

The evaluation was performed on the “score” values for the DDSM database. Table 6.13 below shows the results obtained from a randomly selected experiment.

Table 6.13: SVM score analysis for DDSM database

	Leave One Out		70%/30% Train/ Test		50%/50% Train/Test		10%/ 90% Train/ Test	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
μ	0.973	1.563	0.978	1.314	1.129	1.299	0.742	1.177
σ^2	0.765	1.825	0.698	1.784	0.877	2.133	1.082	2.008
σ	0.586	3.332	0.487	3.182	0.769	4.549	1.172	4.033

Our analysis of SVM score values of the DDSM database leads us to the same conclusions we drew for the Mini-MIAS database earlier. The standard deviation of score values for Normal mammogram cases are relatively small, while for abnormal mammogram cases, the standard deviation of score values are relatively high. This indicates that the mammogram images of the abnormal class are spread over a wider range compared to normal class. Hence in most experiments, the normal mammogram cases can be classified more accurately compared with abnormal mammogram cases.

6.4 Comparison with Existing Works

We now present a comparison of our proposed methods with several existing methods. Table 6.14 and Table 6.15 presents a summary of a comparison between our proposed schemes and various existing works on classifying mammograms as normal or abnormal cases.

Table 6.14: Comparison between proposed method and some related work (Mini-MIAS database)

Methods	Database	Cases	Features	Classificati	Protocol	Results
(Eltoukhy, et al., 2012)	MIAS	322	Multiresolution Statistical Based Features	SVM	70%/ 30% training/testing	94.79%
(Deepa & Bharathi, 2013)	Mini-MIAS	200	Contourlet transform, co-occurrence matrices	PNN	60%/ 40% training/testing	92.5%
(Beura, et al., 2015)	MIAS	322	2D-DWT and GLCM matrix	BPNN	70%/15%/15% training/testing/validation	98.0%
(Caroline & Vaijayanthi, 2016)	Mini-MIAS	317	Multidimensional Complete Ensemble Empirical Mode Decomposition with Adaptive	ANN	52%/48% training/testing	96.7%
Our methods	Mini-MIAS	322	LBP over O+LBP over W-ALL	Linear SVM	Leave-one-out	98.76%
					70%/30% training/testing	98.20%
					50%/50% training/testing	97.69%
					10%/90%training/testing	96.35%
Our methods	Mini-MIAS	322	$f1$	SVM	Leave-one-out	98.45%
					70%/30% training/testing	97.47%
					50%/50% training/testing	97.53%
					10%/90%training/testing	94.17%
Our methods	Mini-MIAS	322	$f2$	SVM	Leave-one-out	96.27%
					70%/30% training/testing	96.13%
					50%/50% training/testing	95.96%
					10%/90%training/testing	91.26%
					Leave-one-out	99.07%

Methods	Database	Cases	Features	Classificati	Protocol	Results
Our methods	Mini-MIAS	322	f_3	SVM	70%/30% training/testing	97.84%
					50%/50% training/testing	97.99%
					10%/90%training/testing	94.83%
Our methods	Mini-MIAS	322	f_4	SVM	Leave-one-out	98.76%
					70%/30% training/testing	97.99%
					50%/50% training/testing	98.21%
					10%/90%training/testing	96.02%
Our methods	Mini-MIAS	322	f_5	SVM	Leave-one-out	96.27%
					70%/30% training/testing	96.60%
					50%/50% training/testing	96.14%
					10%/90%training/testing	91.60%
Our methods	Mini-MIAS	322	f_6	SVM	Leave-one-out	95.96%
					70%/30% training/testing	96.34%
					50%/50% training/testing	95.74%
					10%/90%training/testing	89.55%
Our methods	Mini-MIAS	322	f_7	SVM	Leave-one-out	98.45%
					70%/30% training/testing	97.17%
					50%/50% training/testing	97.16%
					10%/90%training/testing	90.97%

Table 6.15: Comparison between proposed method and some related work (DDSM database)

Methods	Databas	Case	Feature	Classification	Protocol	Results
(Lahmiri & Boukadoum, 2011)	DDSM	100	DWT and Gabor filter bank	SVM polynomial kernel	10-fold cross-validation	98%
(Hussain, et al., 2014)	DDSM	1024	Gabor filter bank	SVM Gaussian kernel	10-fold cross-validation	94.92±23
(Beura, et al., 2015)	DDSM	550	2D-DWT and GLCM matrix	BPNN	70%/15%/15% training/testing/validation	98.8%
Our methods	DDSM	512	LBP over O+LBP over W-ALL	SVM	Leave-one-out	91.80%
					70%/30% training/testing	91.40%
					50%/50% training/testing	90.29%
					10%/90% training/testing	86.00%
Our methods	DDSM	512	$f1$	SVM	Leave-one-out	94.53%
					70%/30% training/testing	95.78%
					50%/50% training/testing	95.10%
					10%/90% training/testing	84.02%
Our methods	DDSM	512	$f2$	SVM	Leave-one-out	87.50%
					70%/30% training/testing	86.85%
					50%/50% training/testing	85.41%
					10%/90% training/testing	78.11%
					Leave-one-out	95.51%

Methods	Databas	Case	Feature	Classification	Protocol	Results
Our methods	DDSM	512	f_3	SVM	70%/30% training/testing	96.04%
					50%/50% training/testing	95.33%
					10%/90%training/testing	85.82%
Our methods	DDSM	512	f_4	SVM	Leave-one-out	95.31%
					70%/30% training/testing	96.20%
					50%/50% training/testing	95.18%
					10%/90%training/testing	86.24%
Our methods	DDSM	512	f_5	SVM	Leave-one-out	85.74%
					70%/30% training/testing	83.31%
					50%/50% training/testing	80.96%
					10%/90%training/testing	75.75%
Our methods	DDSM	512	f_6	SVM	Leave-one-out	84.38%
					70%/30% training/testing	82.99%
					50%/50% training/testing	80.29%
					10%/90%training/testing	74.62%
Our methods	DDSM	512	f_7	SVM	Leave-one-out	91.60%
					70%/30% training/testing	91.49%
					50%/50% training/testing	89.98%
					10%/90%training/testing	77.89%

Before comparing our results with those published, it is important to highlight some differences between the various studies:

- Sample images selected from the Mammogram databases for evaluation might not be the same across all studies.
- Experiment setup differs in different studies. For example, the number of training and testing samples used for experiments are different.
- The number of samples used for the two different classes is not always from the different studies.
- A number of different classes in the database are not equal.

In (Lahmiri & Boukadoum, 2011), authors used Gabor filters and discrete wavelet transform (DWT) for mass detection in the mammogram. They applied Gabor filter banks at different frequencies and spatial orientations of high frequency sub-and images obtained using DWT, then statistical features such as mean and standard deviation are extracted from the Gabor images. SVM with a polynomial kernel was used to classify mammograms as normal or cancer. Their method was tested on just 100 mammograms selected randomly from DDSM database. The accuracy of the system is 98% which is good, but they used only a small sample of the mammograms. In the next chapter, we will introduce the use of super resolution to enhance the mammograms prior to feature extraction. With super-resolved mammograms, we are able to make a small improvement over work of (Lahmiri & Boukadoum, 2011).

Eltoukhy et al. (Eltoukhy, et al., 2012) presented two different functions, (i) distinguish between normal and abnormal tissues, and (ii) determine whether the abnormal tissue is benign or malignant. Firstly, each of the mammograms is decomposed using wavelet transform. Secondly, a set of the coefficients of each mammogram is extracted based on the weight of the feature in discriminating between the different classes. Finally, SVM is used to distinguish between normal and abnormal cases. The accuracy of the system is 94.79% by using Mammographic Image Analysis Society (MIAS) dataset.

In (Deepa & Bharathi, 2013), the authors decompose the ROI using contourlet transform and the co-occurrence matrices in four different directions $\theta = (0^\circ, 45^\circ, 90^\circ \text{ and } 135^\circ)$ and at distance of 1 pixel. A variety of second order statistical texture features (e.g. autocorrelation, Contrast, correlation, cluster prominence, cluster shade,

dissimilarity, energy) are extracted from each co-occurrence matrix. The dimensionality of the features is reduced using Sequential Floating Forward Selection (SFFS) algorithm. A Probabilistic Neural Network (PNN) is used for the classification. A set of 200 mammograms from Mini-MIAS database is used for the evaluation. The accuracy of their system is 92.5%.

In (Hussain, et al., 2014) the authors aimed to classify mammograms into normal and mass regions, and then classify the mass region as benign or malignant. They used Gabor filter banks to extract the textural properties of mammograms at different orientations (3, 5, 6, and 8) and scales (2, 3, 4, and 5). Testing was performed four ROI resolutions (64×64 , 128×128 , 256×256 , and 512×512). They used SVM classifier with Gaussian kernel for the classification. Experiments were conducted using 1024 (512 masses and 512 normal) ROIs extracted from DDSM database. The accuracy of classifying mammograms into normal and mass is 94.92 ± 2.30 .

In (Beura, et al., 2015), the authors aimed to classify the mammogram images to normal, benign and malignant breast tissue. They used GLCM to all the detailed coefficients from 2D-DWT of the region of interest (ROI) of a mammogram and used both t-test and f-test to select the relevant feature from the feature matrix. They used Back Propagation Neural Networks (BPNN) to classify (normal vs. abnormal) and (benign vs. malignant). Experiments were conducted using 322 mammograms from MIAS database 550 mammograms from DDSM database. The accuracy of classifying mammograms to normal or abnormal was 98.0%, and to benign or malignant was 94.2% for MIAS database, and the accuracy of classifying mammograms to normal and abnormal was 98.8% and to benign or malignant was 97.4% for DDSM database.

In (Caroline & Vaijyanthi, 2016), the authors proposed a method to classify the mammogram images to normal, benign and malignant mass. The features were extracted based on Multidimensional Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (MCEEMDAN) which decomposes a given image into frequency components. The ROI was manually cropped to 128×128 pixels for 317 mammograms from Mini-MIAS database. An ANN was used for the classification process. Three experiments were conducted (i) the accuracy of classifying normal and abnormal ROI is 96.7% (ii) classification accuracy of benign and malignant class in abnormal ROI is 96.2% (iii) the accuracy of classifying normal, benign and malignant classes is 96.7%. **Our work has shown that fusion of multiple**

feature types yield much improved performance, and the highest accuracy was achieved with decision level fusion of three feature types (f_3) based schemes. For the Mini-MIAS database, our system had a sensitivity of 98.230%, specificity of 99.522%, and overall accuracy of 99.068%. The sensitivity, specificity and overall accuracy of the DDSM database are 94.531%, 96.484%, and 95.508% respectively.

As can be seen, our proposed methods have improved the classification accuracy over a number of existing methods despite the fact that the performance of all existing works depends not only on the various extracted features but also on sophisticated and expensive learning-based classification such as PNN, BPNN, and ANN. This obvious reason for these results comes from using fusion, and in particular decision level of fusion. For example, the (f_3) scheme is based on the decision level fusion of the LBP over O+LBP over W_ALL, PCA-HOG, and GLCM each of which singularly have good discriminating characteristics and performance. Another important factor in achieving such excellent performance is that of combining the decision based fusion of the three different feature schemes (LBP, GLCM, and HOG) with dimensionality reduction. The added benefits from reducing (balancing) feature vector size before fusing (at decision or feature level) was shown in Chapter 5. In fact, the PCA dimension reduction was necessary for large feature vector size such as HOG descriptor. Moreover, the results obtained from using decision level fusion strategy of balanced feature vector sizes has better classification accuracy when compared to using a single or to a feature level fusion, specially for high performing features.

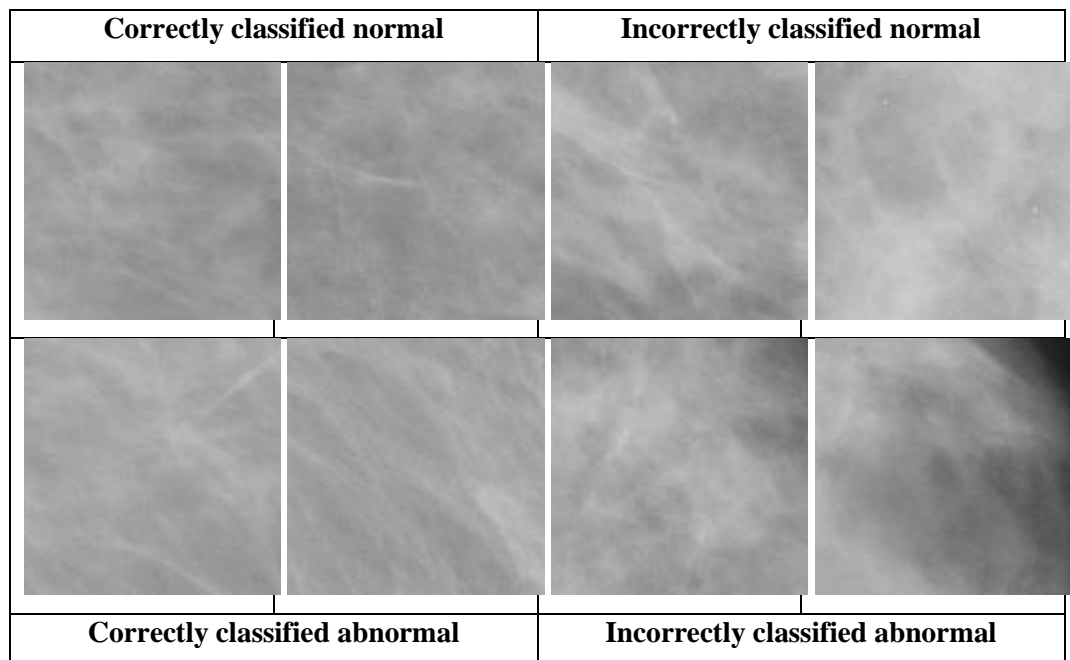
6.5 Conclusions

This chapter proposed feature and decision fusion to improve mammogram classification accuracy. In feature fusion, two or more feature vectors from the same mammogram were concatenated into a single vector and entered to a single classifier. In decision level fusion, majority voting was used to arrive at a final classification based on the classification results of three individual feature sets obtained from the same mammogram. From the statistical evaluation, we conclude that the normal cases are more correctly classified than the abnormal cases because the abnormal mammogram cases include a wide range of abnormalities inside the database (e.g. calcifications; well-defined, spiculated or ill-defined masses; architectural distortion or asymmetry). Also, the severity of each abnormality is different (benign or malignant).

After we evaluated proposed techniques using two databases with different evaluation protocols, we conclude that fusing feature sets of vastly different sizes has negative effects on the classification accuracy when compared to fusing feature sets of similar or small sizes. In other words, fusing of balanced feature set sizes will produce better classification accuracy because of this reason we use a method to reduce the size of the features.

In the decision level fusion strategy better classification accuracy has been obtained when compared with using single or fused features especially when we fuse different feature with near feature vector size.

Finally, we conclude that using decision level fusion of our proposed multi-scale LBP, applying PCA over LBP or HOG, and GLCM features will result in higher classification accuracy compared other single or feature fusing strategies. The classification accuracy for Mini-MIAS database are 99.068%, 97.990%, 98.210%, and 96.017% for Leave-one-out, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing strategies, respectively. While for DDSM database, the classification accuracies are 95.507%, 96.201%, 95.332%, and 86.239% for Leave-one-out, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing strategies, respectively. Figure 6.5 shows examples of correctly and incorrectly classified mammograms when of PCA-HOG and PCA-LBP features were used as fused features.



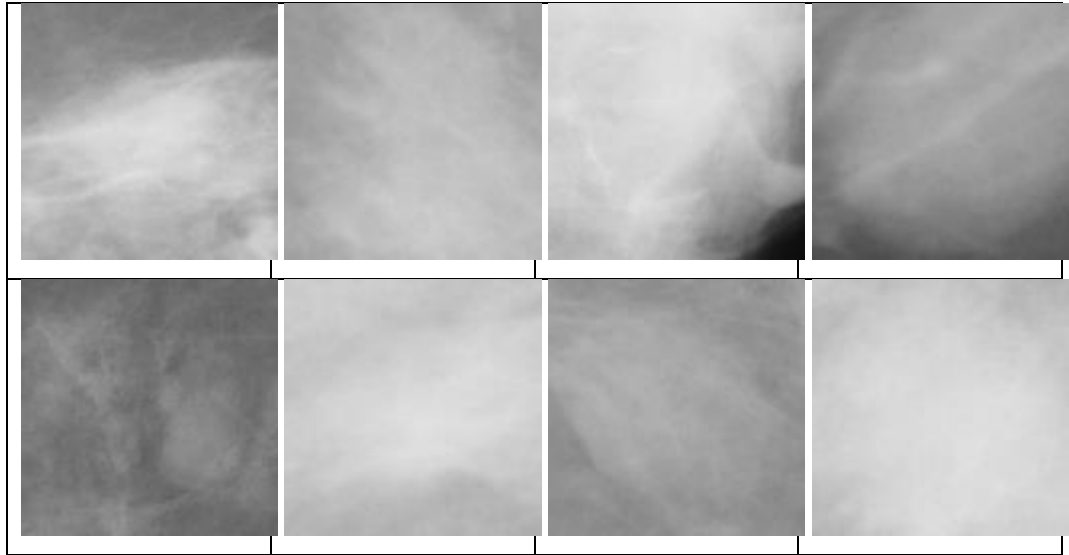


Figure 6.5: Examples of Correctly and Incorrectly classified mammograms

In the next chapter, we shall investigate the use of super resolution to enhance the original mammograms of the two databases prior to feature extraction and classification.

Chapter 7

Mammogram Enhancement using Super resolution

The quality and the resolution of the input images are important factors that affect the overall classification accuracy of automatic image classification systems, including mammogram classification. Image resolution is related to the amount of details available in the specific image area which is measured by the number of sampled pixels and its representation in the computer memory. Low resolution images could affect the appearance of the abnormality regions in mammograms. Image quality is degraded by two different factors: external and internal. Examples of the external factors are lighting conditions and environmental conditions that may lead to loss of contrast, the addition of noise and blurriness, while the internal factors that degrade the quality of the images include the quality of the various camera components, focus, and motion.

To obtain high resolution mammograms from the mammography device, we need large doses of radiation, which may have harmful effects on patients (Zheng, et al., 2010). To overcome this problem we need to obtain high resolution mammograms from low resolution mammograms without increasing in radiation. Recent research in computer vision proposed methods for obtaining higher resolution images from their low resolution version. Super resolution is a method that reproduces high resolution images from there low resolution versions. Recent studies have shown the successful use of image super resolution in applications such as medical imaging (e.g. computed tomography and magnetic resonance imaging) (El Hakimi & Wesarg, 2013), surveillance systems with CCTV (Ahmad & Li, 2012), and satellite-imaging applications (Zhang, et al., 2014).

In (Pak, et al., 2016), authors presented a new method for breast cancer detection and classification in digital mammogram depends on Non-subsampled Contourlet Transform (NSCT) and super resolution algorithm based on fuzzy learning as a pre-processing step. Several features are extracted such as (compactness, fractal, central

moment, area, eccentricity, spread, and average gray level (AGL), and skewness of each feature is calculated. The AdaBoost algorithm was used for classification. They used MIAS database for the experiments, and the mean accuracy is 91.43% and 6.42% as FPR.

Motivated by the studies on image super resolution, we propose to use image super-resolution to enhance the low resolution mammograms prior to extracting texture features. More specifically, we shall use the recently developed super-resolution technique by (Al-Hassan, et al., 2013) (Al-Hassan, et al., 2013) (Al-Hassan, 2014) which they used to enhance low resolution face images for the purposed of face recognition from a distance. We chose this particular super resolution technique as it does not rely on domain specific training data to super-resolve images. We will demonstrate that computed performance parameters (sensitivity, specificity, and accuracy) from obtained SR mammograms compared to its LR mammogram versions for most cases have better classification accuracy.

A brief description of technique we used to super-resolve low resolution mammogram images is given in Section 7.1. Experimental results to evaluate the effectiveness of the proposed use of super-resolution are presented in Section 7.2.

7.1 Super Resolution Technique

The super resolution technique refers to a process of obtaining higher resolution images for lower resolution ones. The super resolution scheme development by Nadia et al. (Al-Hassan, et al., 2013) aims to construct over-complete dictionaries to use in reconstructing a super resolved image from a single input low resolution image for face recognition at a distance. The scheme has two main stages; the first stage prepares the two compressive sensing (CS) dictionaries (low D_L and high D_H) of different sizes and the second stage super resolves a single input LR image to produce an HR image using the two dictionaries.

In the first stage, the class of over-complete dictionaries of size $m \times n$, where $m < n$ are generated. The algorithm starts to construct such matrices by starting with an invertible matrix $m \times m$ as an initial dictionary matrix and iteratively concatenation of selected sets of linearly independent m -column vectors. While maintaining the full spark property after every addition. The dictionary is initialized by the identity matrix,

$$\text{For } i = 1, \dots, \left\lfloor \frac{(n-m)}{m} \right\rfloor, \text{ let } A_{p_i} = \begin{pmatrix} 1 & \frac{1}{p_i} & \frac{1}{p_i^2} & \dots & \frac{1}{p_i^{24}} \\ \frac{1}{p_i} & 1 & \frac{1}{p_i} & \dots & \frac{1}{p_i^{23}} \\ \frac{1}{p_i^2} & \frac{1}{p_i} & 1 & \dots & \frac{1}{p_i^{22}} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{p_i^{24}} & \frac{1}{p_i^{23}} & \frac{1}{p_i^{22}} & \dots & 1 \end{pmatrix} \quad (7.1)$$

$$C = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{12} \\ \frac{1}{2} & 1 & \frac{1}{2} & \dots & \frac{1}{11} \\ \frac{1}{3} & \frac{1}{2} & 1 & \dots & \frac{1}{10} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{12} & \frac{1}{11} & \frac{1}{10} & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{1}{m} & \frac{1}{m-1} & \frac{1}{m-2} & \dots & \frac{1}{m-11} \end{pmatrix} \quad (7.2)$$

Where p_i' are the distinct real numbers >1 . Then, the $m \times n$ dictionary is obtained from the following matrix after normalising its columns using the l_2 -norm:

$$D = [Id, A_{p_1}, A_{p_2}, A_{p_3}, \dots, A_{p_k}]$$

First high-dictionary D_H has been generated by using the integer's numbers $p_i \in [2,20]$ pi where $i = 1,2,3, \dots, 19$, and the low-dictionary D_L created from a Standard Gaussian Random Matrix (SGRM) with zero mean and variance of one and uses l_2 -norm to normalize each column in the dictionaries.

In the second stage, to reconstruct the HR output image, the SR scheme works as follows: The HR image patches recovered by first computing sparse representation of the input low resolution image patch by using a low dictionary and second estimate high resolution image patch by using a high dictionary and sparse representation. Finally, iterative back project method is used to remove possible artifacts from the sparse representation stage and to eliminate the reconstruction errors in the estimated high resolution image (Al-Hassan, et al., 2013), (Al-Hassan, et al., 2013)). (For more details see Chapter 4 of thesis (Al-Hassan, 2014). Figure 7.1 shows some mammogram examples before and after applying the scheme. More examples can be found in Appendix A. We shall demonstrate that this scheme applicable to our mammogram images and also we shall demonstrate the suitability of previous feature extraction schemes when applied on HR mammograms in the next section.

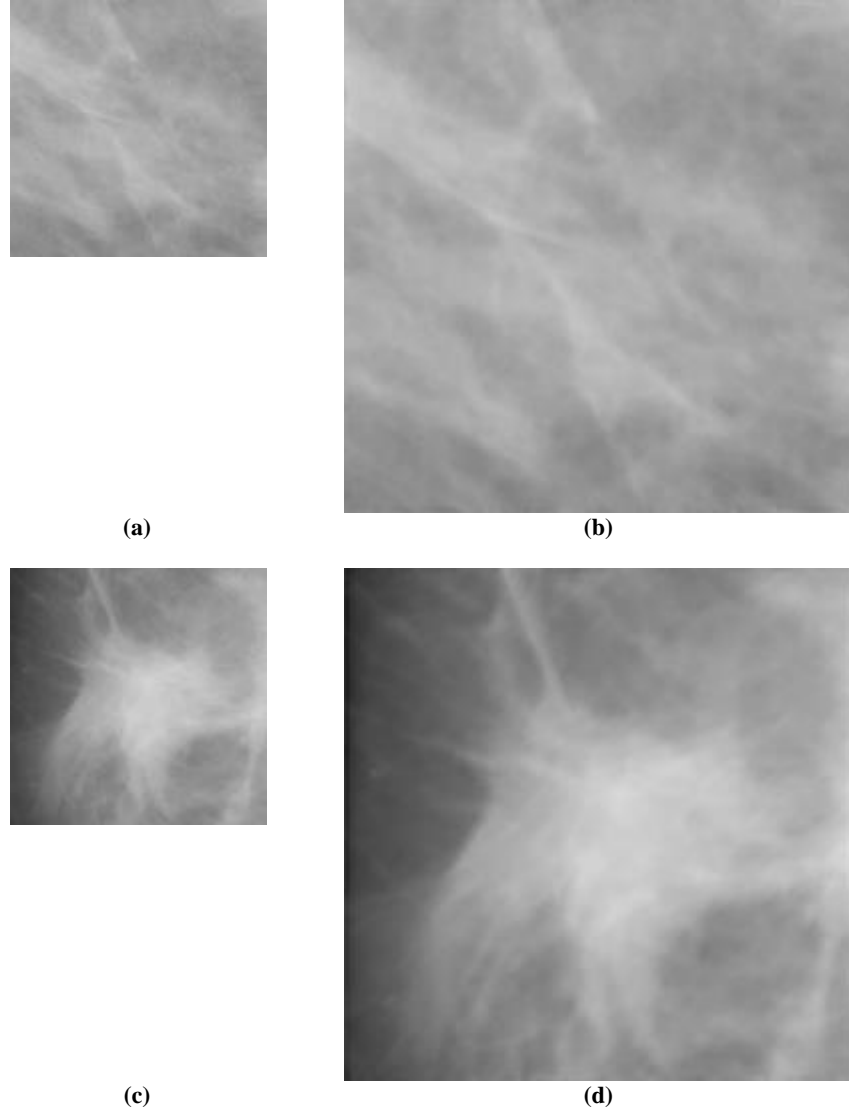


Figure 7.1: An example of Mini-MIAS mammogram: (a) Normal LR-Mini-MIAS mammogram, (b) Normal SR-Mini-MIAS mammogram, (c) Abnormal LR-Mini-MIAS mammogram, and (d) Abnormal SR-Mini-MIAS mammogram

7.2 Experimental Result

We repeated the previous experiments, this time with super-resolved images to evaluate the effectiveness of using SR to enhance the low resolution images in two mammogram databases (Mini-MIAS and DDSM) using the same four strategies for testing and training set sizes (Leave-one-out, 70%/30% training/ testing, 50%/50% training/ testing, and 10%/90% training/testing). Here we used SR mammograms of sizes 256 x 256 with the 8-bit grey value obtained from original LR mammograms of size 128 x 128 using up-mentioned super resolution scheme. As before, the classification was performed using an SVM classifier. The results will be compared with those obtained using the low resolution images for different feature representations, including our

proposed representation, and decision level fusion. In each case, we will measure sensitivity, specificity and accuracy of the system.

7.2.1 Original LBP

Experiments in Chapter 5 showed LBPH features extracted from the original mammogram images resulted in good classification accuracy. Here we repeat the same experiments to evaluate if mammogram enhancement by super resolution yields better classification accuracy. Figure 7.2 and Figure 7.3 show comparative results (sensitivity, specificity, and accuracy) for the original mammograms in Mini-MIAS and DDSM databases and their super-resolved mammograms, respectively based on the four experimental strategies.

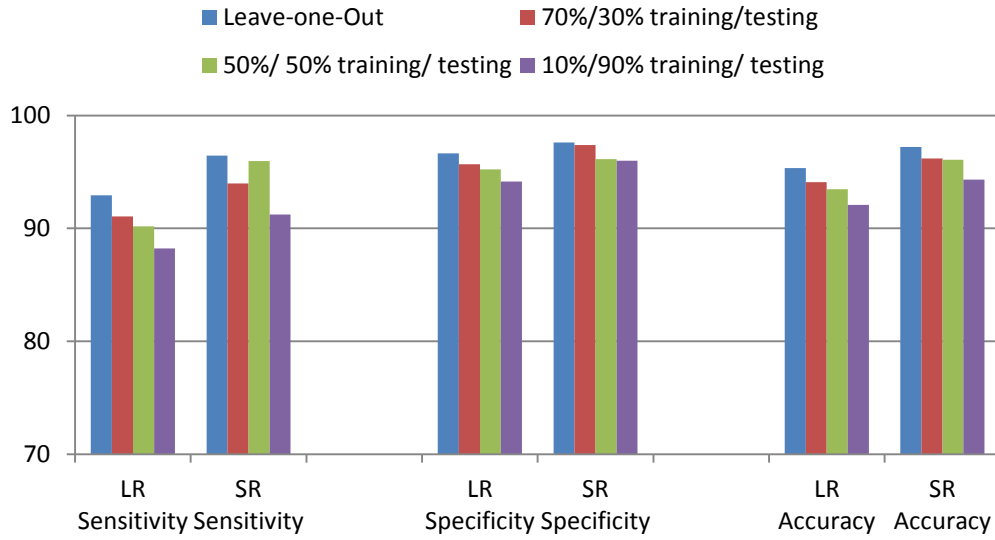


Figure 7.2: Results of Mini-MIAS Database based on LBPH scheme obtained from original mammogram

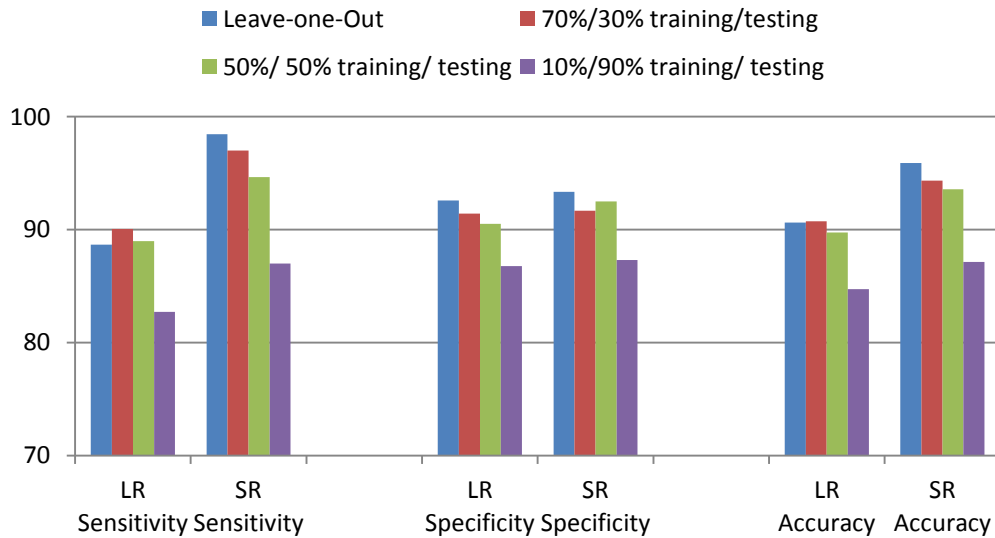


Figure 7.3: Results of DDSM Database based on LBPH scheme obtained from original mammogram

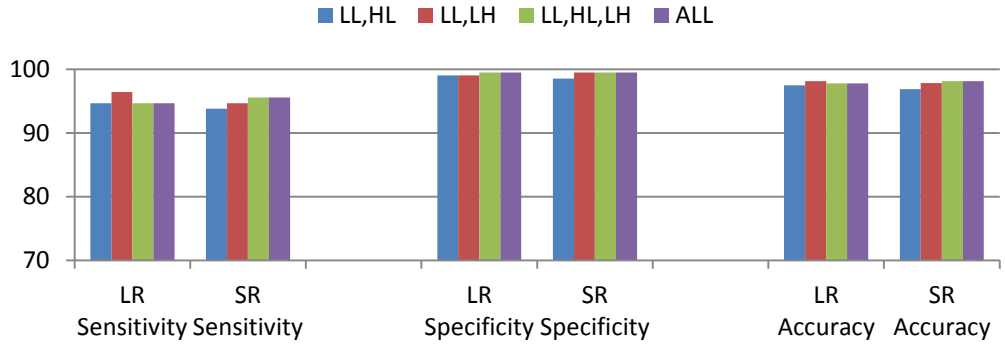
Results show that super resolving the original images of the two databases has led to a significant increase in all three performance indicators. On the Mini-MIAS database, sensitivity improved by 2 to 5 degree; specificity by about 1 to 2 degree; and the overall accuracy by 2 to 3 degree. Meanwhile on the DDMS database, the improvements are about 5 to 10, 1 to 2, and 3 to 5 degree for sensitivity, specificity, and accuracy, respectively. From the classification accuracy we conclude that the LBP feature is good representation of the mammogram image in low resolution and also in super resolution.

7.2.2 Multi Sub-bands LBP

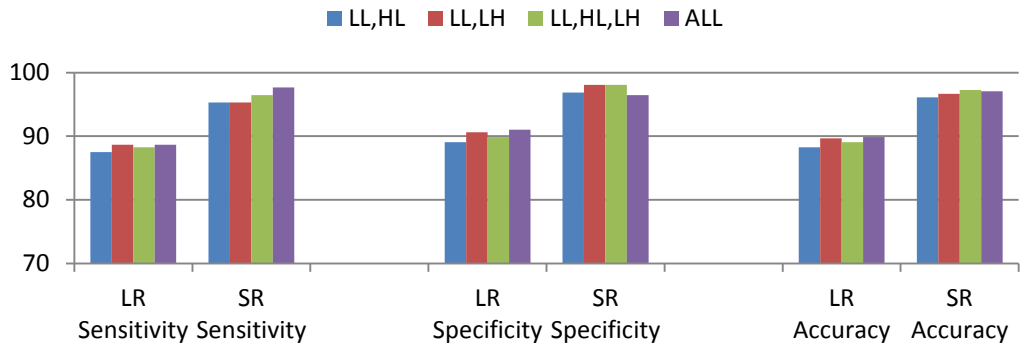
In Chapter 5 we demonstrated that combining LBPH features extracted from multiple wavelet sub-bands provides a better representation of mammogram images than LPBH extracted from an individual sub-bands. Here we repeat the same experiments to evaluate if mammogram enhancement by super resolution yields better classification accuracy. Figure 7.4 to Figure 7.7 shows comparative results for Mini-MIAS and DDSM databases, respectively based on the four experiment strategies.

Results show that super resolving the original images of the two databases has led to a significant increase in all three performance indicators. The performance enhancements follow a very similar pattern when comparing the various databases across all the four experimental strategies.

Because the structure of the breast tissue is very complex and contain mixture of fine lines and dense areas compared to other medical image types, therefore one important advantage of converting the mammogram image to super resolved mammogram is that the detail of the breast is enhanced in a way that make the structure of the breast tissues more clear for classification purpose.

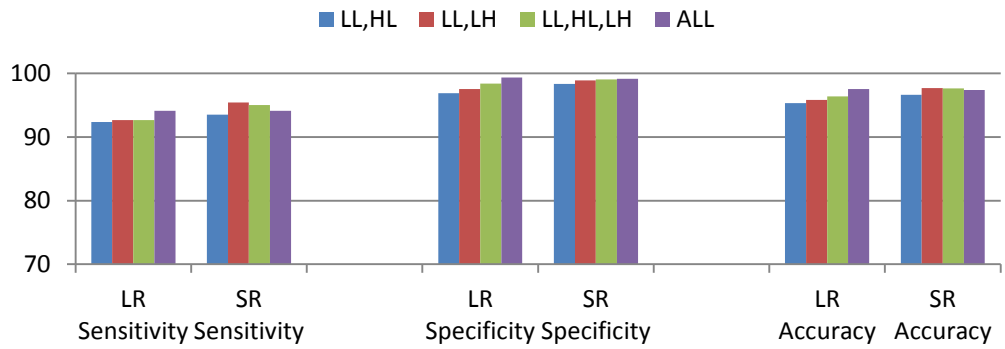


(a)

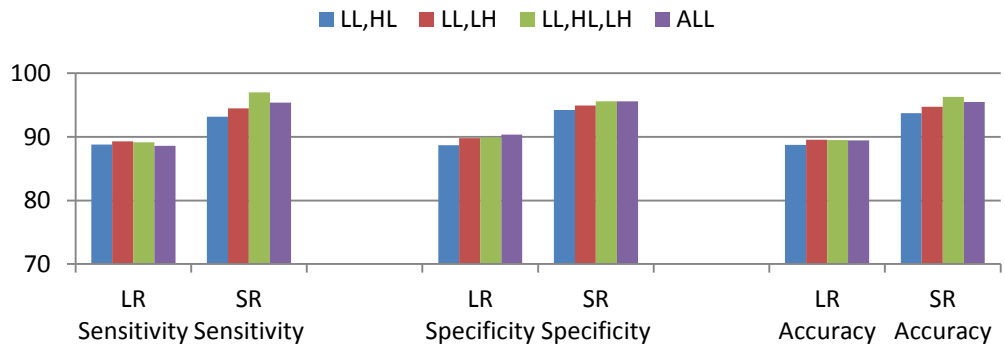


(b)

Figure 7.4: Multi sub-bands results of leave-one-out strategy, (a) Mini-MIAS, (b) DDSM



(a)



(b)

Figure 7.5: Multi sub-bands results of 70% training and 30% testing strategy, (a) Mini-MIAS, (b) DDSM

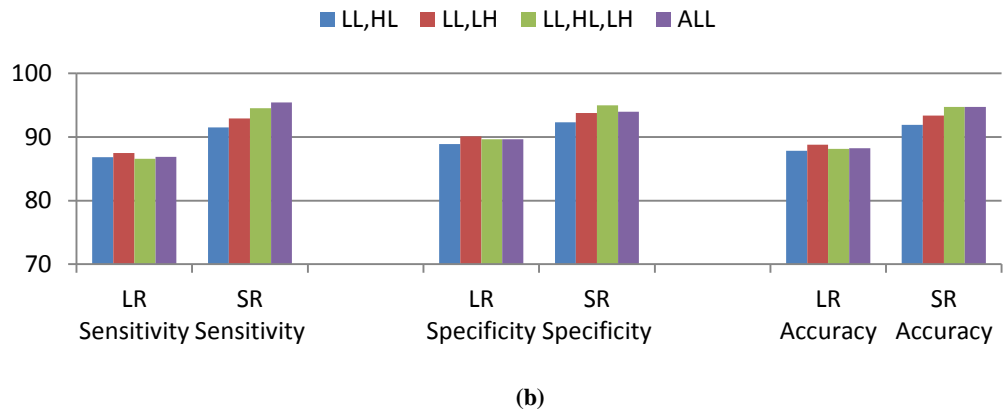
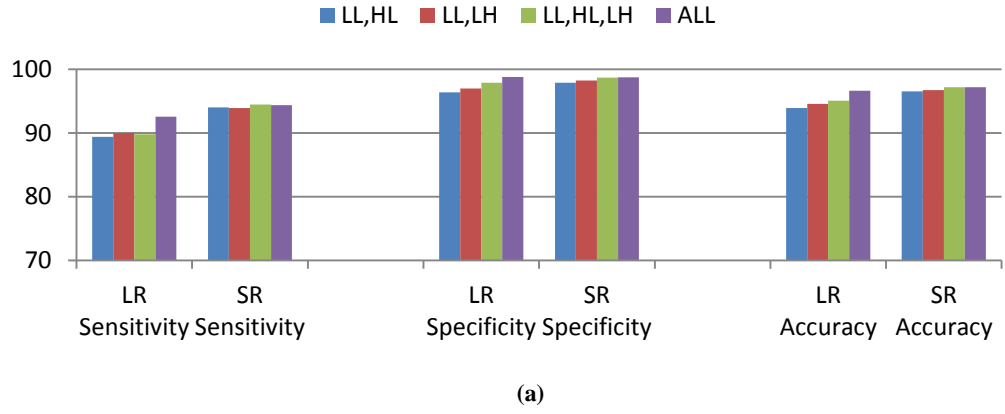


Figure 7.6: Multi sub-bands results of 50% training and 50% testing strategy, (a) Mini-MIAS, (b) DDSM

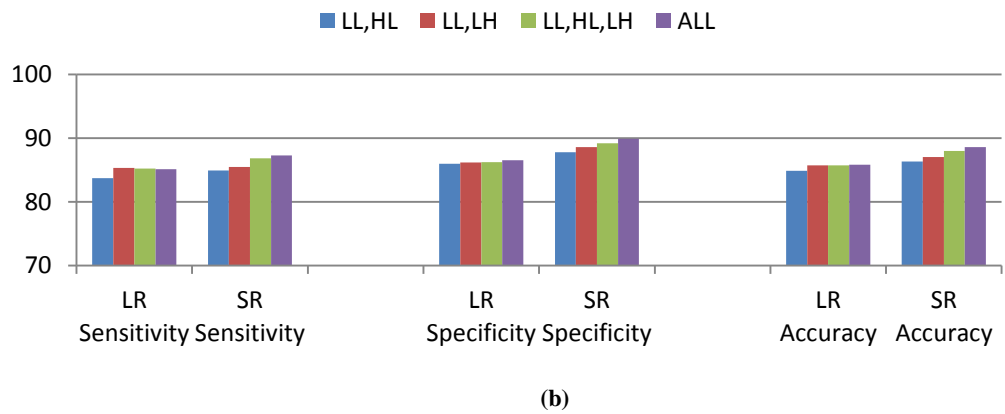
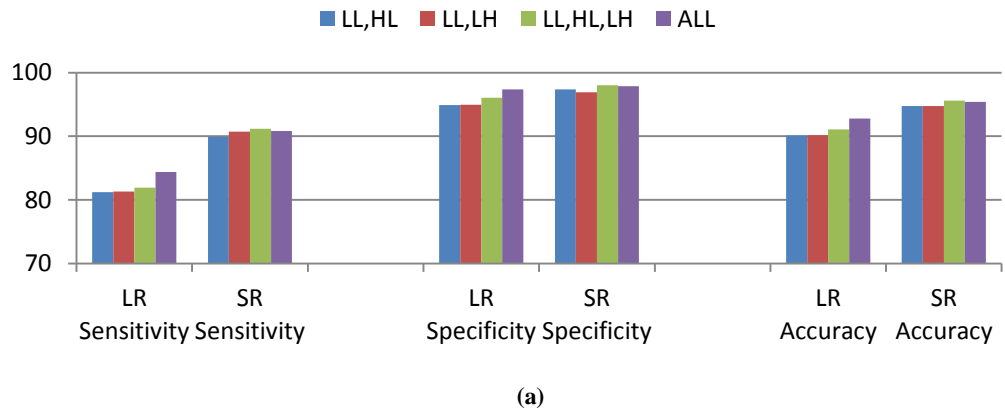
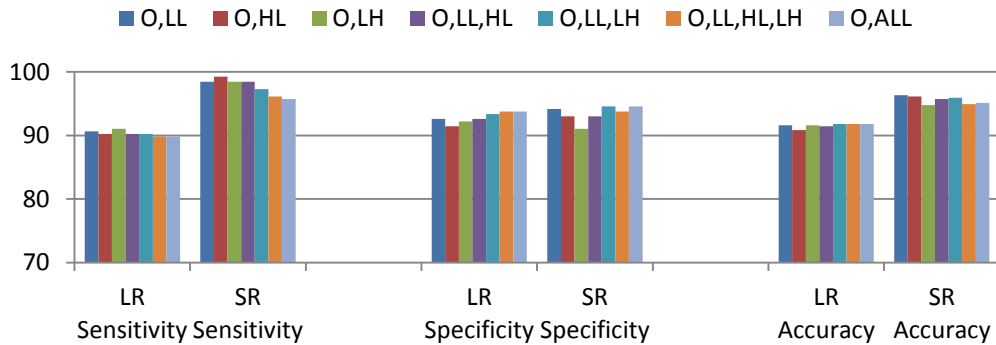
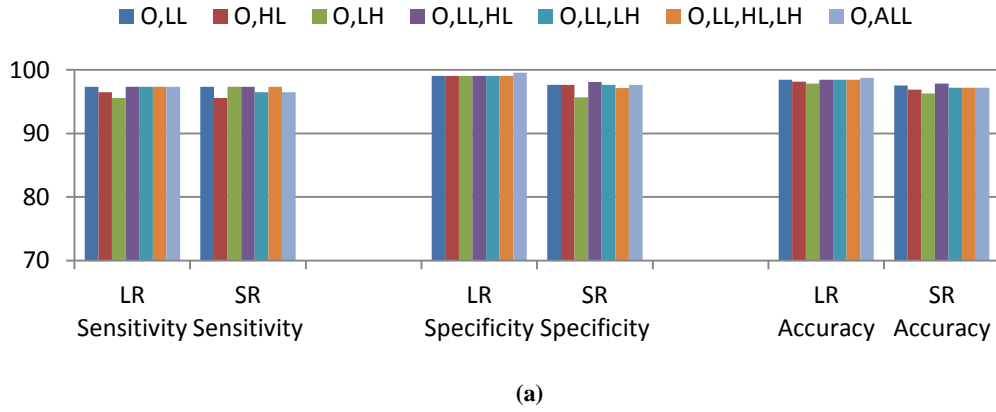


Figure 7.7: Multi sub-bands results of 10% training and 90% testing strategy, (a) Mini-MIAS, (b) DDSM

7.2.3 Original LBP with Multi Sub-bands LBP

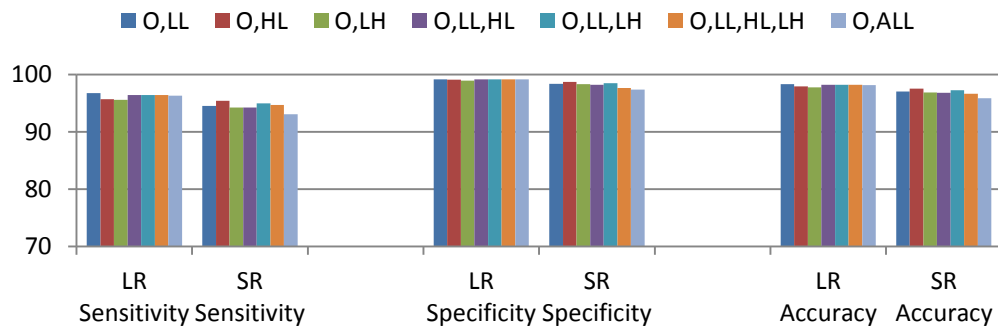
In Chapter 5 we proposed to combine LBPH features extracted from the original spatial domain mammograms and wavelet sub-bands and demonstrated that proposed scheme achieved higher classification accuracy than using only LBPH features extracted from the original mammogram image alone or LBPH features extracted from wavelet sub-bands. This is due to the fact that extracted features from mammogram images are generally thin line features whereby one can contemplate and highlight them in the high frequency wavelet sub-bands. It is worth to mention that using individual wavelet sub-bands did not result in high performance . As mentioned in chapter 5, to overcome this issue, we fused each one of the previously mentioned sub-bands with the original one and this resulted in increasing the performance because adding up more than one sub-band to the original will include features in all directions together with conclusive performance.

Here we repeat the same experiments to evaluate if enhancing mammograms using super resolution yields better classification accuracy. Figure 7.8 to Figure 7.11 shows comparative results for Mini-MIAS and DDSM databases, respectively based on the four experiment strategies.

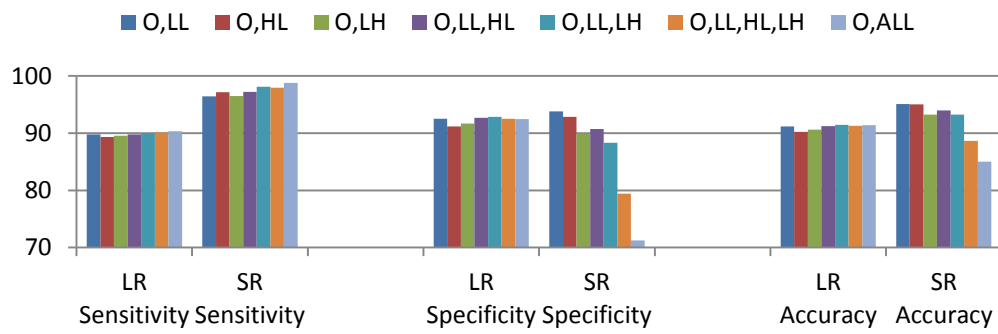


(b)

Figure 7.8: Original LBP and multi sub-bands LBP results of leave-one-out strategy, (a) Mini-MIAS, (b) DDSM

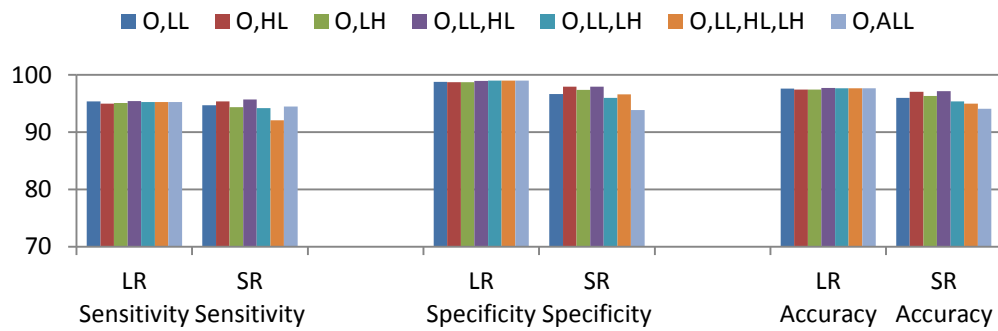


(a)

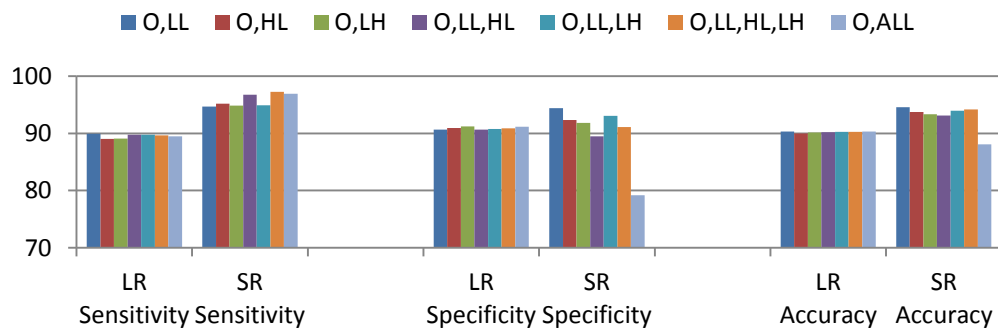


(b)

Figure 7.9: Original LBP and multi sub-bands LBP results of 70% / 30% strategy, (a) Mini-MIAS, (b) DDSM

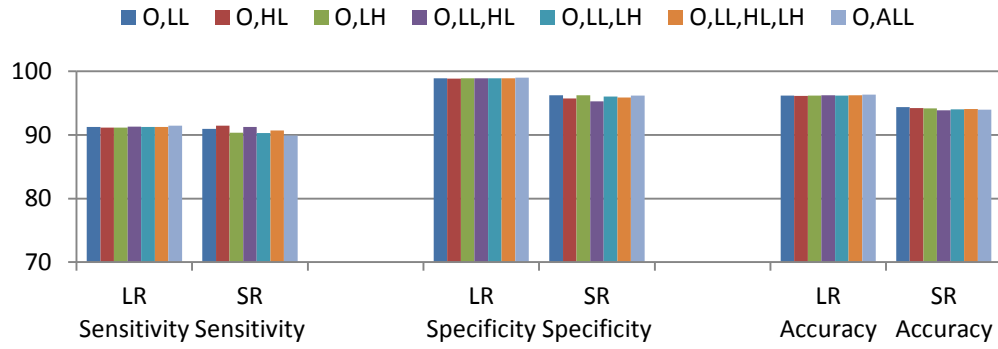


(a)

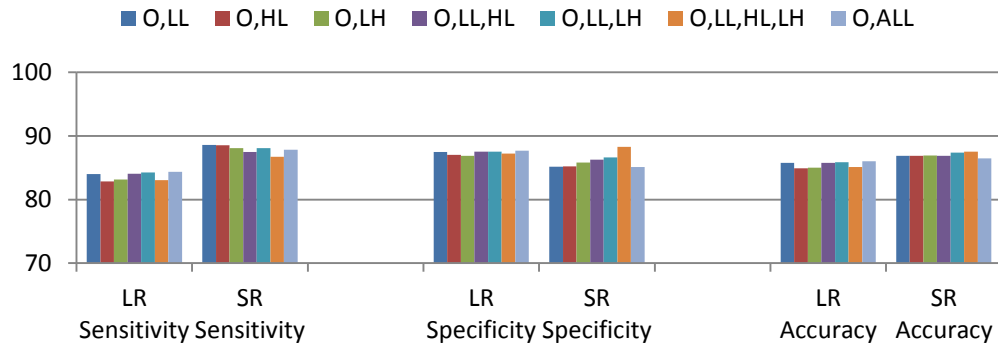


(b)

Figure 7.10: Original LBP and multi sub-bands LBP results of 50% / 50% strategy, (a) Mini-MIAS, (b) DDSM



(a)



(b)

Figure 7.11: Original LBP and multi sub-bands LBP results of 10% / 90% strategy, (a) Mini-MIAS, (b) DDSM

Analysis of the results shows a complex pattern that can be summarised by the following observations:

- For the Mini-MIAS database, SR led to a decrease in overall accuracy. Furthermore, if we look at the individual performance parameters, we can say that decrements occurred more in specificity than in sensitivity.
- For the DDSM database, the patterns are quite the opposite, i.e. in most cases, the overall classification accuracy increased with SR. The increments in accuracy were due mainly to increased sensitivity of the system. This means using SR mammograms made the classification more sensitive than using LR mammograms.
- The above observations apply to all four different experiments strategies we used.
- The type of the mammogram in each database affected the accuracy of the system after convert it to super resolved mammogram, because the structre of the mammogram is different. For example Mini-MIAS database contan more than one

type of abnormal but the DDSM database are just one type (mass) and also they are different in the type of benign and malignant.

7.2.4 HOG Descriptor

We repeated the experiments in Chapter 5 Section 5.2.3 where HOG descriptors were used as mammogram features. Figure 7.12 and Figure 7.13 show comparative results based on SR and LR mammograms for Mini-MIAS and DDSM databases, respectively.

From the results we can say that enhancing mammograms using SR had a positive effect on both databases. Once again, the improvements are prominent for sensitivity than the specificity of the system.

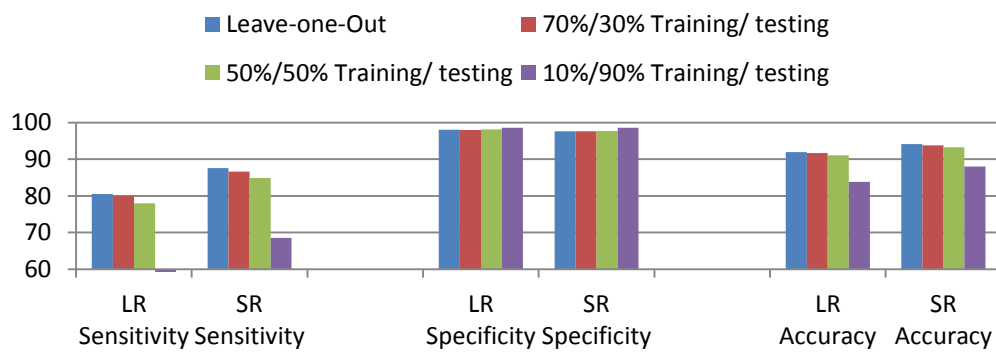


Figure 7.12: Results of HOG features obtained from Mini-MIAS Database

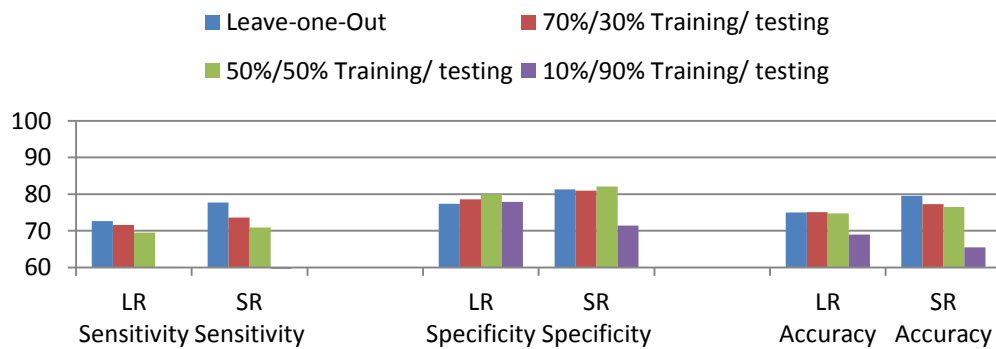


Figure 7.13: Results of HOG features obtained from DDSM Database

Here also HOG features produced the lowest classification accuracy if we compared with all other used feature vectors. A big reason for that could be the feature vector size that represents the mammograms, the information of the mammogram image is different from other types of images, so this affect the accuracy of the system. also we did not consider mass objects types and shapes which might be different from one abnormal mammogram to another and we took all abnormal cases as one type in the experement.

7.2.5 PCA-HOG and PCA-LBP

Experiments in Chapter 5 showed that the large feature sets we proposed for mammogram classification could be reduced to a very small set of PCA based features without compromising classification accuracy. We repeated the same experiments (PCA over original feature vector extracted from LBP and HOG descriptor) to evaluate if mammogram enhancement by super resolution yields better classification accuracy. Feature size was selected based on the highest results achieved with low resolution mammograms; for PCA-LBP we selected PCA feature vector of size 40 and 200 for Mini-MIAS and DDSM respectively, while for PCA-HOG, we used 10 PCA components for Mini-MIAS and 20 PCA components for DDSM. Figure 7.14 and Figure 7.15 shows comparative results for two databases, respectively based on the four experiment strategies.

As we concluded in Chapter 5 reducing number of features affected the classification accuracy in positive way, here we conclude the same thing.

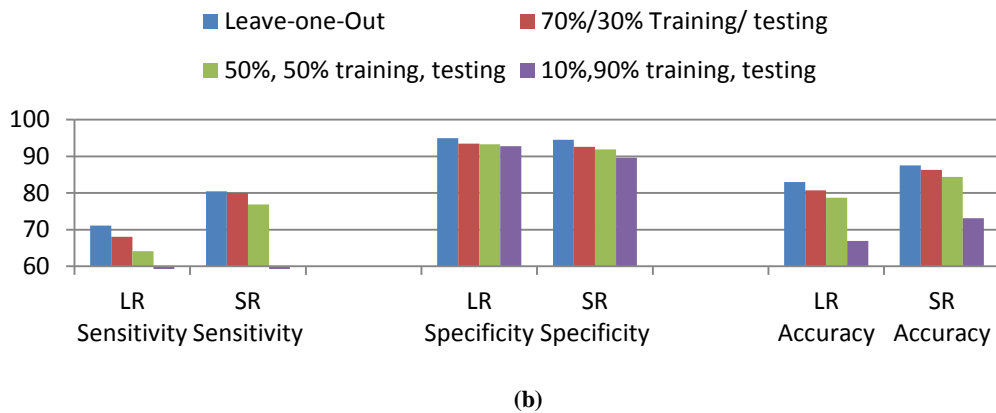
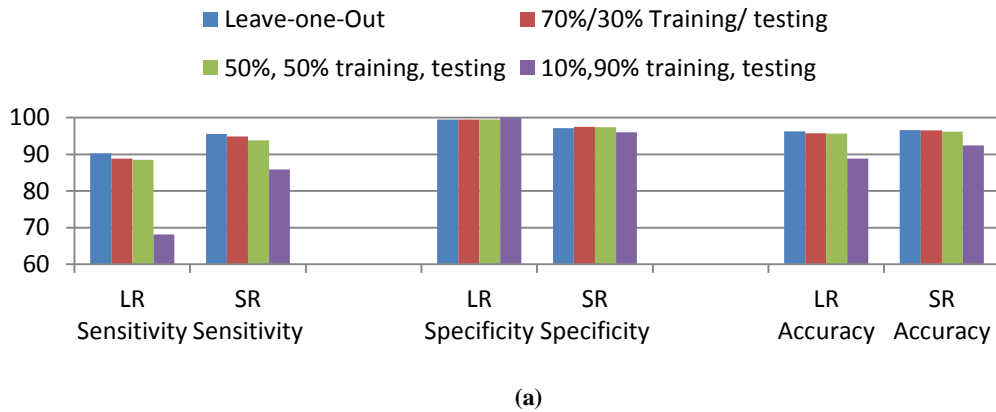
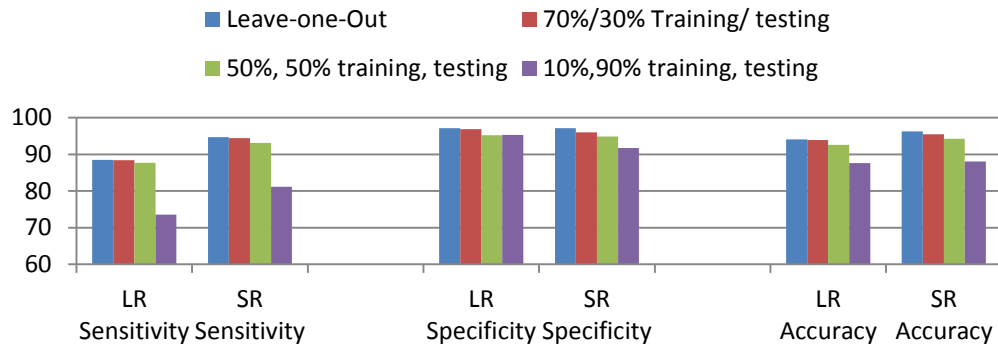
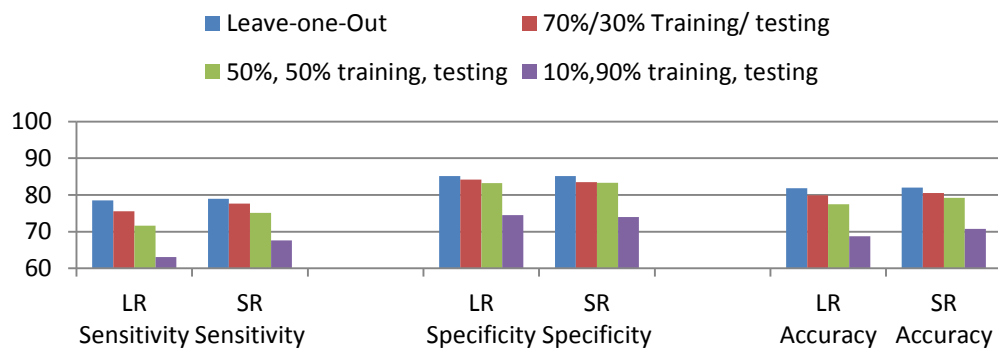


Figure 7.14: Results based on PCA-LBP features obtained from (a) Mini-MIAS, (b) DDSM Database



(a)



(b)

Figure 7.15: Results based on PCA-HOG features obtained from (a) Mini-MIAS, (b) DDSM Database

A close analysis of the results show that super resolving the original images of the two databases has led to:

- A reasonable increase in sensitivity, but a small decrease in specificity
- A small increase in overall classification accuracy, especially when a small number of mammograms are used for training purposes.
- The accuracy is increased by 1 to 4 degrees for Mini-MIAS database, while for DDSM database, the accuracy increased by around 1 to 6 degrees.

7.2.6 GLCM

We repeated the same experiments with GLCM features explained in Chapter 4 and 5 Section 5.2.2. Figure 7.16 and Figure 7.17 shows comparative results for two databases. Super-resolving mammograms in Mini-MIAS database resulted in a decrease in sensitivity, which had a negative effect on the overall classification accuracy. On the other hand super-resolving had a positive effect on the mammograms for DDSM database increased performance in all three measurements. This due to the fact of the

difference between the two databases. From the classification accuracy we conclude that the GLCM feature is not good for all types of mammogram because the Mini-MIAS database contains different abnormal type.

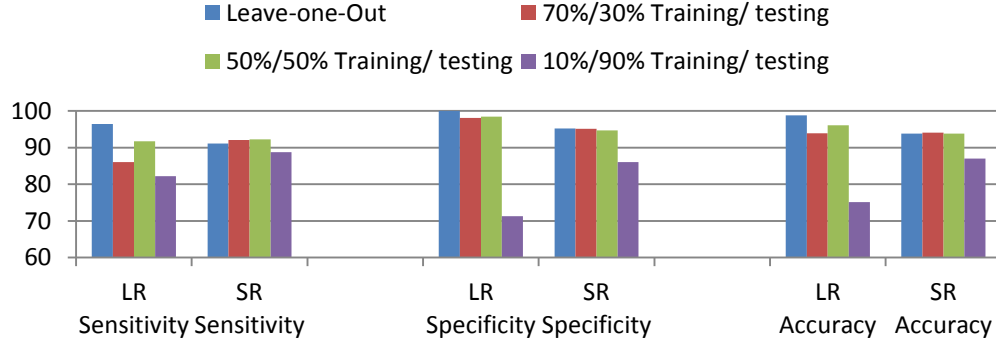


Figure 7.16: Results of GLCM features obtained from Mini-MIAS database

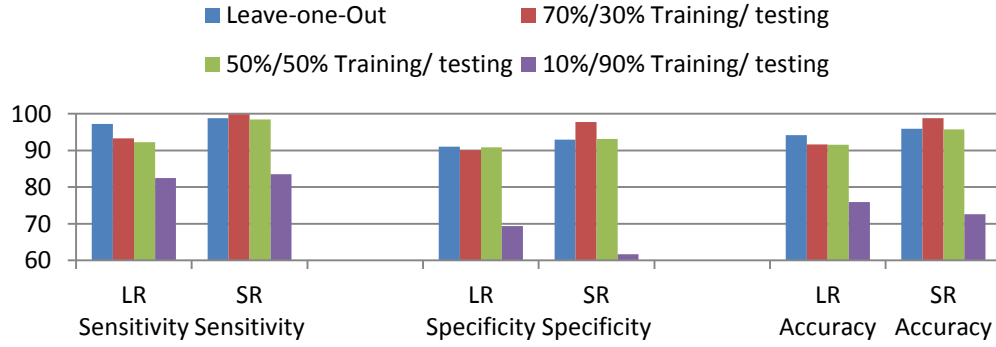


Figure 7.17: Results of using four strategies based on GLCM features obtained from DDSM database

7.2.7 Decision level fusion

We investigated feature fusion and decision fusion in Chapter 6 and proposed decision fusion to enhance classification accuracy. For simplicity we assign letters $f1$ to $f7$ for the 7 different schemes as follows (more details can be found in Chapter 6 Section 6.2.2):

$f1$: fusion of LBP, HOG, and GLCM.

$f2$: fusion of LBP, HOG, and PCA-LBP+PCA-HOG

$f3$: fusion of LBP over O+LBP over W_ALL, PCA-HOG, and GLCM

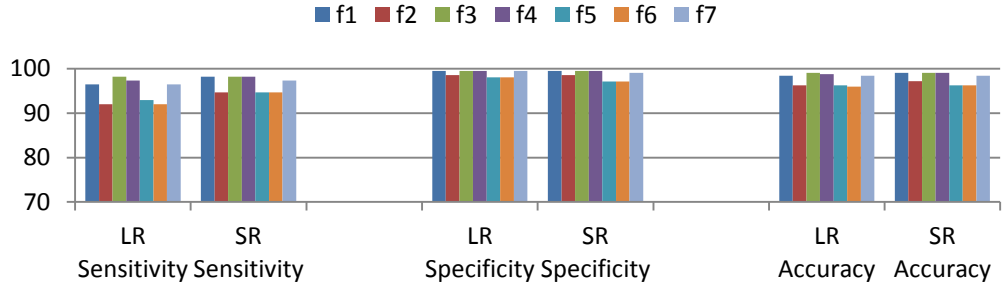
$f4$: fusion of LBP over O+LBP over W_ALL, PCA-LBP+PCA-HOG, and GLCM

$f5$: fusion of LBP over O+LBP over W_ALL, PCA-HOG, and PCA-LBP+PCA-HOG

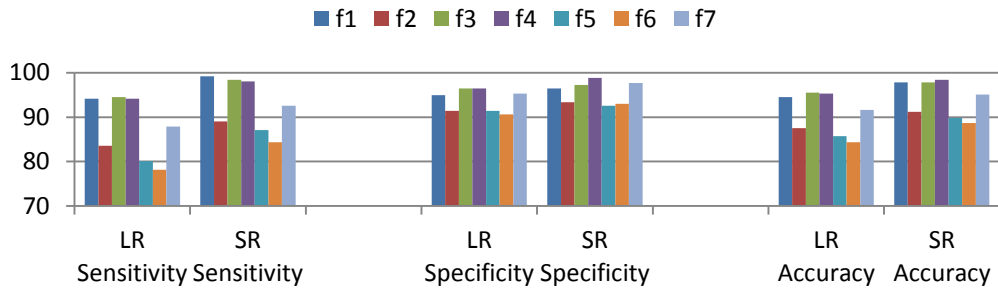
$f6$: fusion of PCA-HOG, PCA-LBP, and PCA-LBP+PCA-HOG

$f7$: fusion of PCA-HOG, PCA-LBP, and GLCM

We repeated the experiments in Chapter 6 with super-resolved mammograms. Figure 7.18 Figure 7.18 to Figure 7.21 shows comparative results for Mini-MIAS and DDSM databases, respectively based on the four experiment strategies.

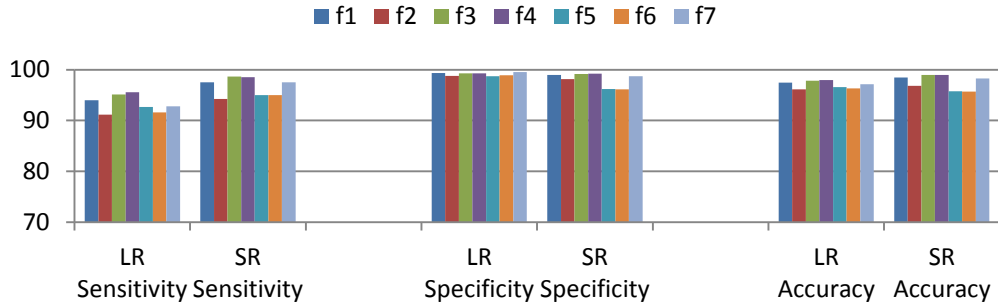


(a)

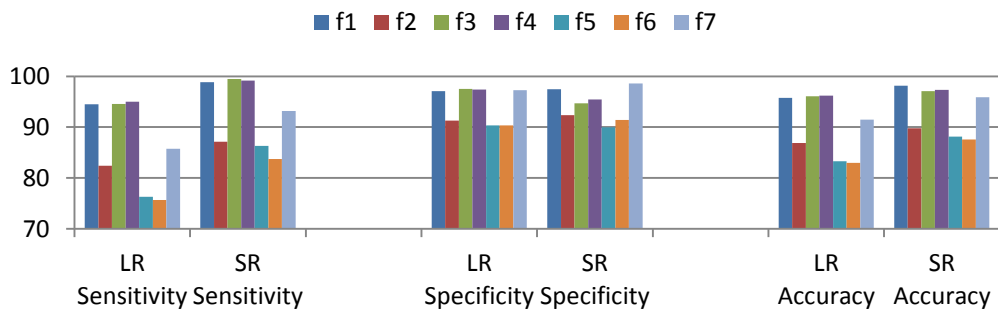


(b)

Figure 7.18: Decision fusion results of Leave-one-out strategy, (a) Mini-MIAS, (b) DDSM



(a)



(b)

Figure 7.19: Decision fusion results of 70% training and 30% testing strategy, (a) Mini-MIAS, (b) DDSM

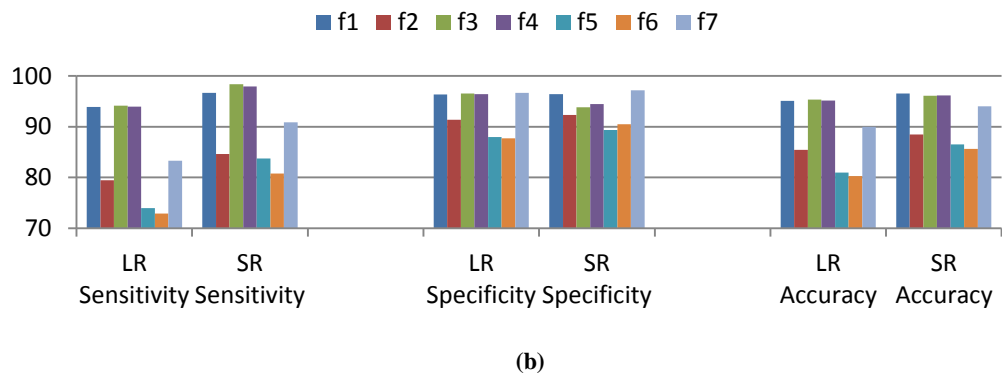
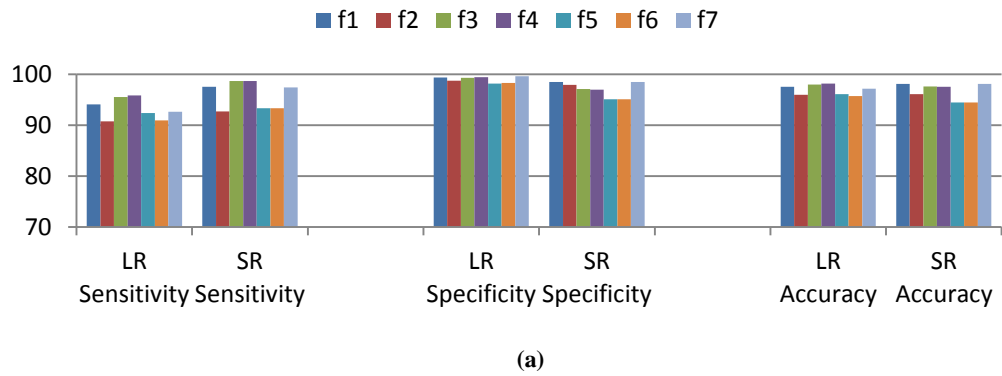


Figure 7.20: Decision fusion results of 50% training and 50% testing strategy, (a) Mini-MIAS, (b) DDSM

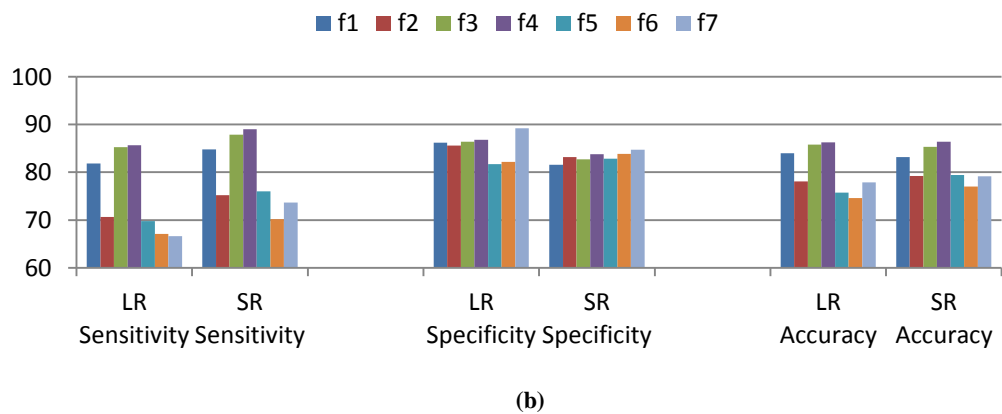
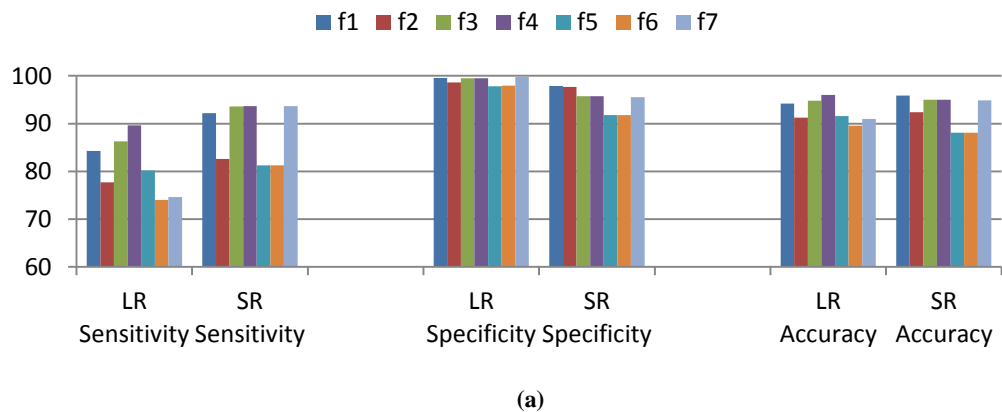


Figure 7.21: Decision fusion results of 10% training and 90% testing strategy, (a) Mini-MIAS, (b) DDSM

Results show that super resolving the original images of the two databases led to an overall increase in performance. Enhancing mammograms using SR led to an increase in sensitivity and the overall accuracy. The specificity also increased in some situations, but decreased on others. Detailed information for individual features can be found in Appendices B and C.

7.3 Conclusions

Recent research in image processing and pattern recognition has demonstrated the effectiveness of image super-resolution in enhancing images resulting in improved classification accuracy in many different applications. Motivated by recent works, we investigated the use of super resolution to enhance mammograms prior to their classification as normal or abnormal. More specifically, we proposed to use a recently developed super resolution scheme originally designed to enhance face images captured at a distance for person identification to enhance mammogram images of the two commonly available mammogram databases (i.e. Mini-MIAS and DDSM).

We repeated experiments conducted in Chapters 5, and 6 to evaluate if super resolving mammogram images could improve system performance in terms of sensitivity, specificity and accuracy. The performance of the system (sensitivity, specificity, and accuracy) was used to compare the performance of original data/images with super-resolved mammogram images or their region of interest.

Analysis of several experimental results showed super-resolving mammograms could achieve a better classification accuracy. More specifically, results show a clear enhancement in terms of sensitivity of the various systems (i.e. reduced False Positives meaning a reduction mammograms wrongly classified as cancer cases). This enhancement was quite prominent on the DDSM database. This due to the facts that using super resolution scheme helps highlighting more textures inside mammograms. In fact, the SR seem to yield improved the discriminating power of all the texture-based features. Moreover, textures are available more in abnormal cases than normal cases, therefore the texture structure of the mammograms are affected (enhanced) by super resolution which leads to recognizing the abnormal cases more accurately.

Although the use of SR led to improved performance in general, it is worth noting that there were instances where SR had a negative effect on performance, particularly on system specificity. The results were not consistent across the two databases. Some

normal cases are wrongly classified as abnormal cases because the super resolution scheme is adding more details to what maybe noise pixels in normal mammograms which can be confused with abnormal cases especially in dense tissue type.

Also it is very important to demonstrate that our proposed method works well with different types of images. From the experiment we noticed that the type of the feature affected the result when the database was converted to super-resolved mammograms. This is likely the case because we have different types of abnormalities in the databases.

We acknowledge that a deeper analysis is required to fully understand and appreciate the effects of super-resolving mammograms on their classification accuracy. This paves the way for further investigation on the use of SR to enhance mammograms. Such future work might require an appropriate dataset. Moreover, one could consider the use of SR to enhance mammograms with microcalcification.

Notwithstanding the limitations highlighted above, this study is, to best of our knowledge, is the first to consider the use of super resolution to enhance mammograms and analysis its effect on classification accuracy.

Chapter 8

Conclusions and Future Work

Breast cancer is one of the major causes of death among women. This thesis was devoted to the problem of automatic detection and diagnosis of breast cancer using mammograms. These investigations were motivated by the widely accepted knowledge that digital mammography screening programs will help to early detection and diagnose of breast cancer and reduces the mortality with increasing the chances of complete recovery. Such screening programs produce a large number of mammographic images that have to be interpreted by radiologists. High-throughput computer aided detection and diagnosis (CAD) systems can help radiologists diagnose accurately and efficiently.

Generally, CAD systems consist of the following main stages: Pre-processing (Segmentation and Enhancement); Feature extraction (analysis); and Classification. Initial investigations revealed that for an automatic mammography screening solution to be practical, the mammograms need to be segmented and enhanced. Consequently, our first task was to determine and extract the breast region of the mammogram, which is defined as the first step in any CAD system which is a pre-processing step. For any CAD system this process is very important for a number of reasons but mainly because it reduces the search zone for detecting the abnormalities or the region of interest of the breast without excessive influence from the background content of the mammograms, which typically includes annotations and artifacts that could have an adverse effect on the analysing results. Moreover, in most Medio-Lateral Oblique (MLO) views of the mammogram the pectoral muscle represents a high density region, therefore, it is important to segment out the muscle during the pre-processing stage of an automated mammogram analysis system because the muscle could bias the detection procedures. These requirements together with the interest in reducing search zone to the only region of interest (ROI) has motivated us to propose a technique to detect the breast region and delete the pectoral muscle.

In the border extraction scheme, adaptive thresholding is used to binarise the mammogram. Then the binarised objects are disconnected using morphological erosion and only the largest object (which is the breast region) inside the image will be kept for

further processing. Then we isolated the spiky boundary and smoothened the remaining region by applying morphological dilation. Finally, the binary mask is mapped with the original mammogram to obtain new ROI mammogram including muscles.

We then removed the pectoral muscle by starting from the top row, moving downwards checking individual pixels to see if they are on the border of muscle with breast region or not and save those on the border. The method we proposed is a snake like technique that tracks the visible thin lines which separate the pectoral muscle from the breast tissue. To detect those thin lines automatically, adaptive thresholding have been used relying on the pixel value intensities and then morphological operations used to remove the labels and artefacts. Finally piecewise linear border oif the pectoral border are drawn between the obtained points from the previous steps to delete the pectoral muscle.

The proposed segmentation was evaluated with the help of an expert radiologist as well as using automatic classification (before and after segmentation) demonstrated that the proposed segmentation works well and have better results compared with results obtained from mammograms without segmentation (see Chapter 4). Moreover, the general conclusion is that segmenting the mammogram and using the relevant ROI helps the classification process.

Our investigation into selecting suitable feature extraction techniques started with reviewing existing schemes. We investigated and proposed a different kind of texture feature extraction methods. The main aim of all proposed schemes is to classify mammograms as normal or abnormal. The proposed features were applied on two publicly available databases; Mini-MIAS and DDSM databases, and with four different testing scenarios with respect to a number of images for training and testing images. The four scenarios are; leave-one-out, 70%/30% training/testing, 50%/ 50% training/testing, and 10%/90% training/testing strategy. The classification was done using a linear SVM classifier. After we evaluated proposed features we can conclude the following:

1. Overall accuracy is higher on the Mini-MIAS database than on the DDSM database. There are differences between the two databases, which we summarise as follows: (i) the total number of cases are different: Mini-MIAS database include 322 ROI and we use all images while for DDSM database we use 512 ROI randomly chosen images. (ii) the number of cases in each class is different: the ratio

between normal and abnormal classes are different as we described in Chapter 2 the number of normal cases is near twice the abnormal cases. This imbalance number of cases might have affected the classification accuracy of the system.

2. A higher classification accuracy was achieved when LBPH features from original mammograms are combined with LBPH features from wavelet sub-bands. This is because the obtained feature vectors from this combination will not miss out on any texture features that could be lost in the multi-resolution wavelet decomposition of the mammogram images. This combination result in better representation of the mammogram images, particularly because of the sensitivity of the method that will highlight mammogram image structure (due to the fact that mammogram images have complex fine lines and dense areas if compared with other medical images).
3. With GLCM statistical features produced better results in terms of the specificity of the system compared to the sensitivity of the system, and as a result, produced better accuracy results. GLCM reflects the mammogram directional features, and features in the direction of 45 and 135 degrees which may not be detectable by the multi-scale LBP, In other words GLCM provides added discriminating mammogram representation.
4. Using HOG features to represent mammogram images produced a very large feature set size. This may have negatively influenced the computation efficiency as well as the classification accuracy of the CAD system despite the fact that we considered all abnormal cases as one type. Another reason for the low classification accuracy by HOG feature is probably due to the fact combining all the 14 different measures may result in high redundancy. One may need to apply some feature selection techniques to reduce the redundancy.
5. Overall, we can conclude that different kind of features will give different classification accuracy result.

Based on the above conclusions, we focused on two aspects; feature set size and fusion. We proposed to use PCA to reduce the high dimensional features to much smaller feature representations. We showed how reduction affects the classification accuracy (see Chapter 5). While in fusion, we use the ideas of feature level fusion and decision level fusion (see Chapter 6). After we evaluated the proposed techniques with different databases and different evaluation protocols we conclude the following:

1. PCA can be used to reduce the number of features that represent mammogram images without sacrificing classification accuracy, or in some cases it led to higher classification accuracy (e.g. applying PCA over HOG feature vector). Moreover, the reduction has a positive affect on the efficiency of the classification system, which is desirable for a high throughput system.
2. Fusing balanced feature set sizes produced better classification accuracy. However, fusing features of different sizes have negative effects on the classification accuracy.
3. Results obtained from using decision level fusion strategy has better classification accuracy when compared to using a single or feature level fusion.
4. We conclude that fusing decisions based on our proposed multi-scale LBP, PCA-LBP or PCA-HOG, and GLCM features can obtain higher classification accuracy than using other single or feature fusing strategies.

The classification accuracy for Mini-MIAS database are 99.068%, 97.990%, 98.210%, and 96.017% for Leave-one-out, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing strategies, respectively. For the DDSM database, the classification accuracies are 95.507%, 96.201%, 95.332%, and 86.239% for Leave-one-out, 70%/30% training/testing, 50%/50% training/testing, and 10%/90% training/testing strategies, respectively. As explained at the end of Chapter 6, the improved accuracy can be attributed to a combination of decision level fusion of a number of high discriminating features with dimension reduction for the higher dimensional features.

Based on recent works on image super resolution and they impact on classification accuracy of different application domains, we focused our attention on enhancing mammograms using super resolution prior to extracting features. More specifically, we applied a recently developed super resolution technique for face recognition to enhance our mammograms. In the majority of cases, super-resolving the mammograms had positive effects on classification accuracy. This is mainly because the super resolution technique results more textures inside mammogram images can be highlighted, especially in abnormal cases whereby it helped to more accurately classify the abnormal cases from normal cases. More highlighted textures means more discriminating characteristics for texture based features, although some noise may become confused with real texture. Removing noise can help in the latter case. More specifically, results show a very clear enhancement in terms of sensitivity of the system,

meaning there were less number of wrongly classified mammograms as cancer cases (see Chapter 7). An advantage of the particular super-resolution technique we used is that it does not depend on any training data, hence it should be applicable on other mammogram datasets. However, further investigation is necessary to have an in depth understanding of the impact of super resolution on mammogram classification.

8.1 Limitations and Future Works

Any study could include some limitations and ours is no exception. We acknowledge that there are a number of limitations in this study, which could be addressed through further investigations. The main limitations and there solutions in the future can be summarized as following:

1. All work done in this thesis were evaluated on two commonly used databases, i.e. Mini-MIAS which include only 322 mammogram cases in total and DDSM database with 512 mammogram cases. This is relatively a small number of tests cases and ideally we would like to have evaluated our proposed schemes on much larger datasets. This also highlights the need to evaluate the developed methods using a larger number of digital mammograms databases.
2. An automated system should be able to segment the ROI within the breast area. We segmented out the breast area from the mammogram by proposed a simple and easy method to implement, but we did not continue with the segmentation to isolate ROIs within the breast area. Since our primary focus was on texture features and fusion, we did not pursue the segmentation further. However, our results show that it is far better to extract features from only the ROI as opposed to the entire breast region.
3. Regarding the previous point (no.2), when the system needs to be practical and needs to be used in a hospital for new cases, we need to find a way for extracting the ROI automatically and applicable for the captured mammograms. Because in our case for extracting ROI we used information available in the databases (ground truth as explained in Chapter 5) such as the size of the mammograms, centre of the mammogram, and the row(x) and column (y) positions of the abnormal centre.
4. The main aim of this thesis was to classify mammogram cases to normal or abnormal cases without looking to the abnormal cases for their specific type. We were driven by the conventional wisdom that first step is to classify mammograms as normal or abnormal and then to classify abnormal cases into different classes (e.g. micro-

calcification, bilateral distortion, mass). However, the statistical analysis presented in Chapter 6 showed that it was more likely that an abnormal case is misclassified as a normal than vice versa. This could be because the abnormal class has so many different types, which should not be presented as one class. Our future work will consider mammogram classification as a multiclass problem from the onset as opposed to the two class problem we dealt with normal or abnormal.

5. We applied PCA to reduce feature vector size. In future, we would like to consider feature selection as an alternative approach to dimension reduction. Feature selection aims to select a subset of relevant features that maximize the classification result.
6. We used two different classifiers (i.e. KNN and SVM). A possible extension could be to employ other types of classifiers ((e.g. SVM with the kernel, ANN, and random forest) for the purpose of classifying the mammogram.
7. We used feature fusion and the decision fusion method and from the results we saw that the accuracy of the system improved in the decision fusion ones. In future we would like to use different classifiers in the same system to classify mammograms based on majority voting (i.e. classifier decision).
8. The use of super resolution led to improved performance in general, but there were instances where it had a negative effect on performance, particularly on system specificity. Moreover, the results were not consistent across the two databases. Therefore, further research is required to fully understand and appreciate the effects of super-resolving mammograms on their classification accuracy. Such future work might require an appropriate dataset. Moreover, one could consider the use of SR to enhance mammograms with micro calcification.

REFERENCES

- Ahmad, T. & Li, X. M., 2012. An Integrated Interpolation-based Super Resolution Reconstruction Algorithm for Video Surveillance. *Journal of Communications*, June.7(6).
- Ahonen, T., Hadid, A. & Pietikäinen, M., 2004. Face Recognition with Local Binary Patterns. *Springer Berlin Heidelberg*, Volume 3021, pp. 469-481.
- Akobeng, A. K., 2007. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatrica*, Issue 3, p. 338–341.
- Al-Hassan, N., 2014. *Mathematically Inspired Approaches to Face Recognition in Uncontrolled Conditions-Super Resolution and Compressive Sensing*, Buckingham, UK: University of Buckingham.
- Al-Hassan, N., Jassim, S. A. & Sellaheewa, H., 2013. *Construction of Dictionaries to Reconstruct High-Resolution Images for Face Recognition*. Madrid, Spain, 6th International Conference on Biometrics.
- Al-Hassan, N., Sellaheewa, H. & Jassim, S. A., 2013. *Super Resolution Based Face Recognition Do we need training image set?*. Baltimor/USA, SPIE Proceedings Vol. 8755.
- Al-Jawad, N., 2009. *Exploiting Statistical Properties of Wavelet Coefficients for Image/Video Processing and Analysis Tasks*, Buckingham, United Kingdom: Dphil thesis.
- Al-Jubouri, H., 2015. *Multi Evidence Fusion Scheme for Content-Based Image Retrieval by Clustering Localised Colour and Texture Features*, Buckingham, United Kingdom: University of Buckingham.
- Atrey, P. K., Hossain, M. A., El Saddik, A. & Kankanhalli, M. S., 2010. Multimodal fusion for multimedia analysis: a survey. *Multimedia systems*, Volume 16(6), pp. 345-379.
- Bandyopadhyay, S. K., 2010. Pre-processing of Mammogram Images. 2(11), pp. 6753-6758.
- Berbar, M. A., Reyad, Y. A. & Hussain, M., 2012. *Breast Mass Classification using Statistical and Local Binary Pattern Features*. Montpellier , IEEE, pp. 486-490.

Beura, S., Majhi, B. & Dash, R., 2015. Mammogram Classification Using Two Dimensional Discrete Wavelet Transform and Gray-Level Co-occurrence Matrix for Detection of Breast Cancer. Volume 154, pp. 1-14.

Bingham, E. & Mannila, H., 2001. *Random Projection in Dimensionality Reduction: Applications to Image and Text Data*. USA, international conference on Knowledge discovery and data mining, pp. 245-250.

Bovis, K. & Singh, S., 2000. *Detection of Masses in Mammograms Using Texture Features*. Barcelona , IEEE, pp. 267-270.

Bozek, J., Delac, K. & Grgic, M., september, 2008. *Computer-Aided Detection and Diagnosis of Breast Abnormalities in Digital Mammography*. Zadar, Croatia, IEEE, pp. 45-51.

Brown, J., Bryan, S. & Warren, R., 1996. Mammography screening: An incremental cost effectiveness analysis of double versus single reading of mammograms. *Br. Med. J.*, 9 January, 312(7034), p. 809–812.

Camilus, K. S., Govindan, V. K. & Sathidevi, P. S., 2010. Computer-Aided Identification of the Pectoral Muscle in Digitized Mammograms. *Journal of Digital Imaging*, October, Volume 23, pp. 562-580.

Cancer Research, UK, 2014. *Breast cancer mortality statistics*. [Online] Available at: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/mortality> [Accessed 2015].

Caroline, B. E. & Vijayanthi, N., 2016. Feature Extraction of digital Mammogram Based on Multidimensional Complete Ensemble Empirical Mode Decomposition With Adaptive Noise. *International Journal of Advanced Engineering Technology*, VII(I), pp. 404-409.

Chai, Y. et al., 2015. Hierarchical and Multi-featured Fusion for Effective Gait Recognition under Variable Scenarios. *Pattern Analysis and Applications*, pp. 1-13.

Chai, Y., Wang, Q., Jia, J. & Zhao, R., 2006. A Novel Gait Recognition Method Via Fusing Shape and Kinematics Features. *Advances in Visual Computing*, Volume 4291, pp. 80-89.

- Chandrasekhar, R. & Attikiouzel, Y., 1996. *Gross Segmentation of Mammograms using a Polynomial Model*. Amsterdam, IEEE, pp. 1056-1058.
- Che Kuo, Y., Chen Lin, W., Chang Hsu, S. & Chun Cheng, A., 2014. *Mass Detection in Digital Mammograms System Based on PSO Algorithm*. Taichung, IEEE International Symposium on Computer, Consumer and Control (IS3C), pp. 662-668.
- Chen, Z. & Zwiggelaar, R., 2010. *Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms*. Coventry, UK, s.n., pp. 71-75.
- Christoyianni, I., Koutras, A., Dermatas, E. & Kokkinakis, G., 2002. Computer aided diagnosis of breast cancer in digitized mammograms. *Computerized Medical Imaging and Graphics*, 26(5), pp. 309-319.
- Costaridou, L., 2005. *Medical Image Analysis Methods*. Boca Raton: CRC Press.
- da Fonseca, J. L., February 2013. *Pre-CADs in Breast Cancer*, Porto: University Do Porto.
- Dalal, N. & Triggs, B., 2005. *Histograms of Oriented Gradients for Human Detection*. San Diego, Computer Vision and Pattern Recognition (CVPR).
- Das, P., Bhattacharyya, D., Bandyopadhyay, S. K. & Kim, T.-h., 2009. Analysis and Diagnosis of Breast Cancer. *International Journal of u- and e- Service, Science and Technology*, September.2(3).
- de Cruz, C. F., July, 2011. *Automatic Analysis of Mammography: Enhancement and Segmentation Techniques*, Porto: Engineering Faculty, Porto University.
- Deepa, S. & Bharathi, V. S., 2013. Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN. *Journal of Computer Engineering (IOSR-JCE)*, 10(10), pp. 7-13.
- Déniz, O., Bueno, G., Salido, J. & De la Torre, F., 2011. Face Recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, Volume 32, p. 1598–1603.
- Deserno, T. M., Soiron, M., de Oliveira, J. E. & Araújo, A. A., 2011. *Towards computer-aided diagnostics of screening mammography using content-based image retrieval*. Maceio, Alagoas , IEEE 24th SIBGRAPI Conference on Graphics, Patterns and Images (Sibgrapi).

- Dhawan, A. P., Chitre, Y., Bonasso, C. & Wheeler, K., 1997. Radial-Basis-Function Based Classification of Mammographic Microcalcifications Using Texture Features. *IEEE-EMBC and CMBEC*, pp. 535-536.
- Doi, K., 2007. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. *Comput Med Imaging Graph*, Volume 31, pp. 198-211.
- Dong, A. & Wang, B., 2009. *Feature Selection and Analysis on Mammogram Classification*. Victoria, BC, IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 731-735.
- Duda, R. O., Hart, P. E. & Stork, D. G., November 2000. *Pattern Classification*. Second edition ed. Canada: John Wiley and Sons.
- Eberl, M. M. et al., 2005. BI-RADS Classification for Management of Abnormal Mammograms. *JABFM*, 19(2), pp. 161-164.
- Eddaoudi, F., Regragui, F., Mahmoudi, A. & Lamouri, N., 2011. Masses Detection Using SVM Classifier Based on Textures Analysis. *Applied Mathematical Sciences*, 5(8), pp. 368-379.
- El Hakimi, W. & Wesarg, S., 2013. *Accurate Super-Resolution Reconstruction for CT and MR Images*. Porto, IEEE, pp. 445 - 448.
- Elshinawy, M. Y., 2010. *Pre-CAD Normal Mammogram Detection Algorithm Based on Tissue Type*, Washington, D.C.: HOWARD UNIVERSITY.
- Elshinawy, M. Y., Badawy, A. A., Abdelmageed, W. W. & Chouikha, M. F., 2010. *Pre-CAD System for Normal Mammogram Detection using Local Binary Pattern Features*. Perth, WA, IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 352-357.
- Eltoukhy, M., Faye, I. & Samir, B., 2009. *Breast Cancer Diagnosis in Mammograms Using Multilevel Wavelet Analysis*. Malaysia, National Postgraduate Conference at Universiti Teknologi PETRONAS.
- Eltoukhy, M. M., Faye, I. & Brahim, S., 2012. A Statistical Based Feature Extraction Method for Breast Cancer Diagnosis in Digital Mammogram Using Multiresolution Representation. *Computers in Biology and Medicine*, 42(1), pp. 123 -128.

- Fathima, M. M., Manimegalai, D. & Thaiyalnayaki, S., 2013. *Automatic Detection of Tumor Subtype in Mammograms Based On GLCM and DWT Features using SVM*. Chennai, IEEE International Conference Information Communication and Embedded Systems (ICICES), pp. 809-813.
- Ferrari, R. J. et al., 2004. *Automatic Identification of the Pectoral Muscle in Mammograms*. Canada, IEEE TRANSACTIONS ON MEDICAL IMAGING, pp. 232-245.
- Ferrari, R. J., Rangayyan, R. M., L., J. E. & Frère, A. F., 2000. *Segmentation of mammograms: Identification of the skin – air boundary, pectoral muscle, and fibro-glandular disc*. Canada, 5th International Workshop on Digital Mammography, p. 573–579.
- Field, A., 2006. *Discovering Statistics Using SPSS*. second edition ed. London, UK: SAGE publications.
- Gonzales, R. C. & Woods, R. E., 2002. *Digital Image Processing*. Second eddition ed. New Jersey: Tom Robbins.
- Gunn, S. R., 1998. *Support Vector Machines for Classification and Regression*, UK: Faculty of Engineering, Science and Mathematics.
- Guo, G. et al., 2003. KNN Model-Based Approach in Classification. *Lecture Notes in Computer Science*, Volume 2888, p. 986 – 996.
- Guo, Z., Zhang, L., Zhang, D. & Mou, X., 2010. *Hierarchical Multiscale LBP for Face and Palmprint Recognition*. Hong Kong, International Confrance on Image processing.
- Hanahan, D. & Weinberg, R. A., 2011. Hallmarks of Cancer: The Next Generation. *Cell*, March, pp. 646-674.
- Haralick, R. M., K., S. & Dinstein, I., 1973. Texture Features for Image Classification. *IEEE Transactions on systems, Man and Cybematics*, November, 3(6), pp. 610-621.
- Hashimoto, B. E., 2008. *Practical Digital Mammography*. New York: Thieme Medical Publishers.
- Hastie, T., Tibshirani, R. & Friedman, J., February 2009. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Second ed. s.l.:Springer.

- Heath, M. et al., 2001. *The digital database for screening mammography*. s.l., Medical Physics Fifth International Workshop on Digital Mammography, pp. 212-218.
- Heisele, B., Ho, P. & Poggio, T., 2001. *Face Recognition with Support Vector Machines: Global versus Component-based Approach*. Canada, IEEE International Conference Computer Vision, pp. 1-7.
- Huang, D. et al., 2011. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, November , 41(6), pp. 765-781.
- Hussain, M. et al., 2014. Effective Extraction of Gabor Features for False Positive Reduction and Mass Classification in Mammography. *Applied Mathematics & Information Sciences*, Volume 8(1L), pp. 397-412.
- Jolliffe, I. T., 2002. *Principle Component Analysis*. Second ed. New York: Springer.
- Karahaliou, A. N. et al., 2008. Breast Cancer Diagnosis: Analyzing Texture of Tissue Surrounding Microcalcifications. *IEEE Transactions on Information Technology in Biomedicine*, November, 12(6), pp. 731-738.
- Karssemeijer, N., 1998. Automated classification of parenchymal patterns in mammograms. 43(2), pp. 365-378.
- Khuzi, A. M., Besar, R., Zaki, W. W. & Ahmad, N., 2009. Identification of masses in digital mammogram using gray level co-occurrence matrices. *Biomedical Imaging and Intervention Journal*, June.5(3).
- Kobayashi, T., Akinori Hidaka, A. & Kurita, T., 2008. *Selection of Histograms of Oriented Gradients Features for Pedestrian Detection*. s.l., Springer-Verlag Berlin Heidelberg, pp. 598-607.
- Kong, S., Wang, X., Wang, D. & Wu, F., 2013. *Multiple Feature Fusion for Face Recognition*. China, 10 th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).
- Kramer, D. & Aghdasi, F., 1998. Classification of Microcalcifications in Digitised Mammograms Using Multiscale Statistical Texture Analysis. *IEEE*, pp. 121-126.
- Krishnaveni, S., Bhanumathi, R. & Pugazharasan, T., 2014. Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier. *International*

Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, March, 3(3), pp. 8274-8282.

Kwan¹, M. L. et al., 2009. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. May.11(3).

Kwok, S. M., Chandrasekhar, R., Attikiouzel, Y. & Rickard, M. T., 2004. Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms. *IEEE Transactions on Medical Imaging*, 9, 23(9), pp. 1129-1140.

Lahmiri, S. & Boukadoum, M., 2011. *Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Mammogram Features Extraction*. France, New Circuits and Systems Conference (NEWCAS).

Lee, N. J., December, 2006. *Computer-Aided Diagnostic Systems for Digital Mammograms*, Louisiana: Louisiana State University.

Linguraru, M. G., Marias, K., English, R. & Brady, M., 2006. A Biologically Inspired Algorithm for Microcalcification Cluster Detection. *Medical Image Analysis*, 10(6).

Majeed, T. F., Al-Jawad, N. & Sellaheewa, H., 2013. *Breast border extraction and pectoral muscle removal in MLO mammogram images*. Essex,UK, IEEE 5th Computer Science and Electronic Engineering Conference (CEEC), pp. 119-124.

Mallat, S. G., 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July, 11(7), p. 674–693.

Marti, J., Freixenet, J., Munoz, X. & Oliver, A., 2003. *Active region segmentation of mammographic masses based on texture, contour and shape features*. s.l., Springer-Verlag Berlin Heidelberg, p. 478–485.

Marx, C. et al., 2004. Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of Mammographic diagnosis with and without use of CAD. *European Journal of Radiology*, Volume 51, p. 66–72.

Masala, G. L., 2006. Computer Aided Detection on Mammography. *World Academy of Science, Engineering and Technology*, Volume 15.

Mika, S. et al., 1999. *Kernel PCA and De-Noising in Feature Spaces*. s.l., MIT Press Advances in Neural Information Processing Systems, pp. 536-542.

- Mirzaalian, H., Ahmadzadeh, M. R. & Sadri, S., 2007. *Pectoral Muscle Segmentation on Digital Mammograms by Nonlinear Diffusion Filtering*. Victoria, BC , IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 581-584.
- Mohanty, A. K., Beberta, S. & Lenka, S. K., 2011. Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram. *Engineering Research and Applications (IJERA)*, 1(3), pp. 687-693.
- Molloy, J., 1997. *How Mammography is Performed: Imaging and Positioning*. [Online]
Available at: <http://www.imaginis.com/mammography/how-mammography-is-performed-imaging-and-positioning-2#mediolateral-oblique-view-mlo>
[Accessed 06 November 2014].
- Mustra, M., Bozek, J. & Grgic, M., 2009. *Breast Border Extraction and Pectoral Muscle Detection using Wavelet Decomposition*. St.-Petersburg ,Croatia, IEEE EUROCON 2009, pp. 1428-1435.
- Mustra, M. & Grgic, M., 2013. Robust Automatic Breast and Pectoral Muscle Segmentation from Scanned Mammograms. *Elsevier, SignalProcessing*, 93(10), pp. 2817-2827.
- Nagi, J., Abdul Kareem, S., Nagi, F. & Ahmed, S. K., 2010. *Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms*. Kuala Lumpur, Malaysia, IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES), pp. 87-92.
- Naveed, N., Choi, T.-S. & Jaffar, A., 2011. Malignancy and abnormality detection of mammograms using DWT features and ensembling of classifiers. *International Journal of the Physical Sciences Vol. 6(8)*, pp. 2107-2116, 18 April, 2011, April, 6(8), pp. 2107-2116.
- NIH, U. S. N. I. o. H., September, 2009. *What you need to know about breast cancer*, s.l.: s.n.
- Ojala, T., Pietikäinen, M. & Harwood, D., 1996. A Comparative Study of Texture Measures with Classification based on Featured distributions. *Pattern Recognition*, January, 29(1), pp. 51-59.

- Ojala, T., Pietikainen, M. & Maenppa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July, 24(7), pp. 971-987.
- Oliver, A., Llad, X., Freixenet, J. & Marti, J., 2007. False Positive Reduction in Mammographic Mass Detection Using Local Binary Patterns. *MICCAI, Springer-Verlag Berlin Heidelberg*, 4791(1), p. 286–293.
- Oral, C. & Sezg, H., 2013. *Effects of dimension reduction in mammograms classification*. Bursa , IEEE 8th International Conference on Electrical and Electronics Engineering (ELECO).
- Pak, F., Kanan, H. R. & Alikhassi, A., 2016. Breast Cancer Detection and Classification in Digital Mammography Based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution. *Comput Methods Programs Biomed*, 122(2), pp. 89-107.
- Ponraj, N. D., Sweety, K., Dr., P. P. & J. , S. M., 2011. A Survey on Texture Analysis of Mammogram for the Detection of Breast Cancer. *Digital Image Processing*, 3(15).
- Raba, D. et al., 2005. *Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms*. Berlin Heidelberg, Pattern Recognition and Image Analysis Springer, pp. 471-478.
- Raba, D., Oliver, A., Marti, J. & Perac, M., 2005. *Breast segmentation with pectoral muscle suppression on digital mammograms*. Portugal, Pattern Recognition and Image Analysis, Springer, pp. 471-478.
- Rangayyan, R. M., 2005. *Biomedical Image Analysis*. Boca Raton: CRC Press LLC.
- Rangayyan, R. M., Ayres, F. J. & Desautels, J. L., 2007. A review of computer-aided diagnosis of breast Cancer: Toward the detection of Subtle Signs. *Journal of the Franklin Institute*, May, 334(3-4), pp. 312-348.
- Rashid, R. D., Jassim, S. A. & Sellahewa, H., 2013. *LBP based on multi wavelet sub-bands feature extraction used for face recognition*. Southampton, UK, IEEE.
- Ross, A. & Govindarajan, R., 2005. *Feature Level Fusion Using Hand and Face Biometrics*. USA, SPIE Conference on Biometric Technology for Human Identification.

- Sabir, A., Al-jawad, N. & Jassim, S., 2014. *Feature selection gait-based gender classification under different*. Brussels, Belgium, SPIE, Real-Time Image and Video Processing II.
- Sabu, A. M., Ponraj, D. N. & D., 2012. Textural Features Based Breast Cancer Detection: A Survey. *Journal of Emerging Trends in Computing and Information Sciences*, septamber, 3(9), pp. 1329-1334.
- Sampat, M. P., Gupta, S., Markey, M. K. & Bovik, A. C., 2008. *Computer Aided Detection and Diagnosis in Mammography*, s.l.: s.n.
- Sampat, M. P., Gupta, S., Markey, M. K. & Bovik, A. C., 2008. *Computer Aided Detection and Diagnosis in Mammography*, s.l.: s.n.
- Sample, J. T., 2003. *Computer assisted screening of digital mammogram images*, Louisiana: Louisiana State University.
- Samulski, M., 2006. *Classification of Breast Lesions in Digital Mammograms*, Nijmegen, Netherlands: Radboud University Nijmegen.
- Sellahewa, H., 2006. *Wavelet-based Automatic Face Recognition for Constrained*, Buckingham, United Kingdom: Dphil thesis.
- Sellahewa, H. & Jassim, S., 2010. Image Quality-based Adaptive Face Recognition. *IEEE Transactions on Instrumentation and Measurement*, April, 59(4), pp. 805-813.
- Sellahewa, H. & Jassim, S. A., 2008. *Illumination and expression invariant face recognition: toward sample quality-based adaptive fusion*. Arlington, VA, 2nd IEEE International Conference on Biometrics: Theory, Application and Systems.
- Senthilkumar, B. & Umamaheswari, G., 2011. A Review on Computer Aided Detection and Diagnosis - Towards the Treatment of Breast Cancer. *European Journal of Scientific Research*, 52(4), pp. 437-452.
- Sharma, M., Dubey, R. B., S. & Gupta, S. K., 2012. Feature Extraction of Mammograms. *International Journal of Advanced Computer Research*, September, 2(5), pp. 192-199.
- Shinde, M., 2003. *Computer aided diagnosis in digital mammography: Classification of mass and normal tissue*, Florida: University of South Florida.
- Society, A. C., 2009. *Breast Cancer*, s.l.: American Cancer Society.

- Subramaniam, E., Liung, T. K., Yusoff Mashor, M. & Isa, N. A. M., 2006. Breast Cancer Diagnosis Systems: A Review. *International Journal of The Computer, the Internet and Management*, May - August, 14(2), pp. 24-35.
- Suckling, J. et al., 1994. *The Mammographic Image Analysis Society Digital Mammogram Database*. York, England, Elsevier Science, pp. 375-378.
- Suhas, S. S., Ajay , K. B. & Khanale, P. B., 2012. Face Recognition Using Principal Component Analysis and Linear Discriminant Analysis on Holistic Approach in Facial Images Database. *IOSR Journal of Engineering*, 2(12), pp. 15-23.
- Taher, A., 2014. *Unsupervised Cooperative Partitioning Approach of Hyperspectral Images for Decision Making*, Lannion, France: Universite Rennes.
- Taher, A., Chehdi, K. & Cariou, C., 2014. *An Unsupervised Nonparametric and Cooperative Approach for Classification of Multicomponent Image Contents*. Angers, France, International Conference on Pattern Recognition Applications and Methods (ICPRAM 2014).
- Tang, H., Sun, Y., Yin, B. & Ge, Y., 2010. *Face recognition based on haar LBP histogram*. Chengdu , IEEE 3rd International Conference on Advanced Computer Theory and Engineering(ICA CTE).
- Tang, J., Rangayyan, R. M., El Naqa, I. & Yang, Y., 2009. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*, March, 13(2), pp. 236-251.
- Tan, P.-N., Steinbach, M. & Kumar, V., 2006. *Introduction to Data Mining*. Pearson International Edition ed. s.l.:Addison Wisley.
- The Independent UK Panel on Breast Cancer, S., October 2012. The Benefits and Harms of Breast Cancer Screening: An Independent Review. *The Lancet*, 380(9855), pp. 1778-1786.
- Tzikopoulos, S. D. et al., 2010. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *computer methods and programs in biomedicine*, November, Issue 30, p. 47–63.
- Tzikopoulos, S. D. et al., 2011. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *computer methods and programs in biomedicine Elsevier Ireland*, Issue 02, pp. 47-63.

UK.GOV, 2010. [Online]
 Available at: <http://www.nhs.uk/conditions/Cancer-of-the-breast-female/Pages/Introduction.aspx>
 [Accessed 22 September 2014].

Vaidehi, K. & Subashini, T. S., 2013. Automatic Identification and Elimination of Pectoral Muscle in Digital Mammograms. *International Journal of Computer Applications*, 8, 75(14), pp. 15-18.

Varjonen, M. et al., 2007. Breast positioning system for full field digital mammography and digital breast tomosynthesis system. *Medical Imaging 2007: Physics of Medical Imaging*, Volume 6510.

Vieux, R., Benois-Pineau, J. & Domenger, J.-P., 2012. *Content Based Image Retrieval Using Bag-of-regions*. Berlin, Heidelberg, Springer-Verlag, pp. 507-517.

Vikhe, P. S. & Thool, V. R., 2016. Intensity Based Automatic Boundary Identification of Pectoral Muscle in Mammograms. *7th International Conference on Communication, Computing and Virtualization*, Volume 79, p. 262 – 269.

Winsberg, F. et al., 1967. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, Volume 89, pp. 211-215.

Wirth, M. A. & Stapinski, A., 2003. *Segmentation of the breast region in mammograms using active contours*. Lugano, Switzerland, SPIE Visual Communications and Image Processing .

Xue, Q., Yan, Z. & Wang, S., 2010. *Mammographic Feature Enhancement Based on Second Generation Curvelet Transform*. Yantai , IEEE 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010), pp. 349-353.

Y. Wang, Q. Y. a. K. L., 2011. *Hand Vein Recognition based on Multi-scale LBP and Wavelet*. Guilin, IEEE International Conference on Wavelet Analysis and Pattern Recognition, pp. 214-218.

Yacoub, M. A., Mohamed, A. S. & Kadah, Y. M., 2006. *A CAD System for the Detection of Malignant Tumors in Digitized Mammogram Films*. Egypt, PROC. CAIRO INTERNATIONAL BIOMEDICAL ENGINEERING CONFERENCE.

Yanfeng, L. et al., 2015. Texton analysis for mass classification in mammograms. *Pattern Recognition Letters*, Volume 52, pp. 87-93.

- Youssry, N., Abou-Chadi, F. E. & El-Sayad, A. M., 2003. *Early Detection of Masses In Digitized Mammograms Using Texture Features and Neuro-Fuzzy Model*. UK, IEEE 4th Annual Conference on Information Technology Applications in Biomedicine, pp. 226-229.
- Yu, S. & Guan, L., 2000. A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, FEBRUARY, 19(2), pp. 115-126.
- Yu, Z., Jixian, . Z. & J., V. g. L., 2006. *Comparison and Analysis of Remote Sensing Data Fusion Techniques at Feature and Decision Levels*. s.l., Zeng, Y., Zhang, J. and Van Genderen, J.L., 2006. Comparison and analysis of remote sensing data fusion techniques at fe ISPRS Commission VII Mid-term Symposium; Remote Sensing: From Pixels to Processes.
- Zhang, H., Yang, Z., Zhang, L. & Shen, H., 2014. Super-Resolution Reconstruction for Multi-Angle Remote Sensing Images Considering Resolution Differences. 6(1), pp. 637-657.
- Zhang, Z., Lu, J. & Yip, J., 2008. *Computer Aided Mammography*. Huddersfield, UK, School of Computing and Engineering Researchers' Conference, University of Huddersfield.
- Zhao, G. & Pietikainen, M., 2007. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. 29(6), pp. 1-14.
- Zheng, J., Fuentes, O. & Leung, M.-Y., 2010 . *Super-resolution of Mammograms*. Montreal, QC , Computational Intelligence in Bioinformatics and Computational Biology (CIBCB),IEEE.

APPENDIX A: Some examples of SR-Mini-MIAS database

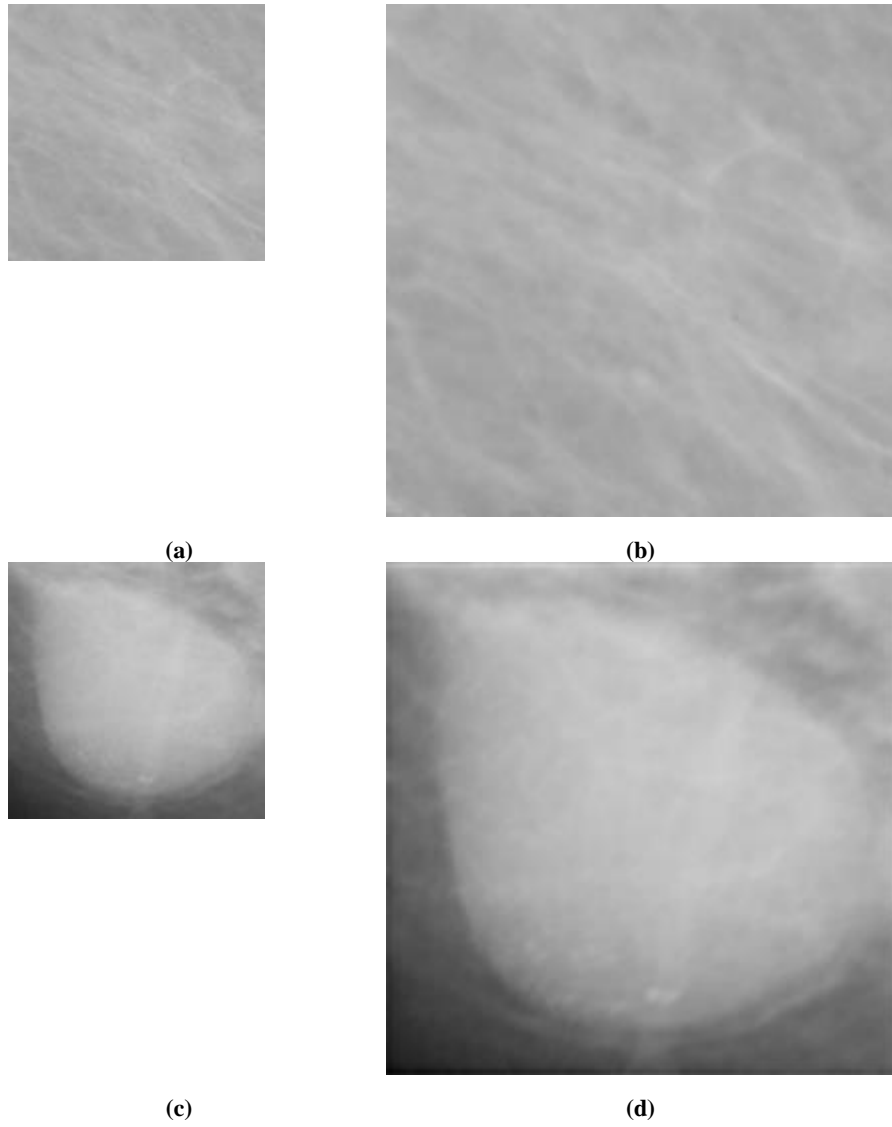


Figure A.1: Example1 for Mini-MIAS mammogram: (a) Normal LR-Mini-MIAS mammogram 128x128, (b) Normal SR-Mini-MIAS mammogram 256x256, (c) Abnormal LR-Mini-MIAS mammogram 128x128, and (d) Abnormal SR-Mini-MIAS mammogram 256x256

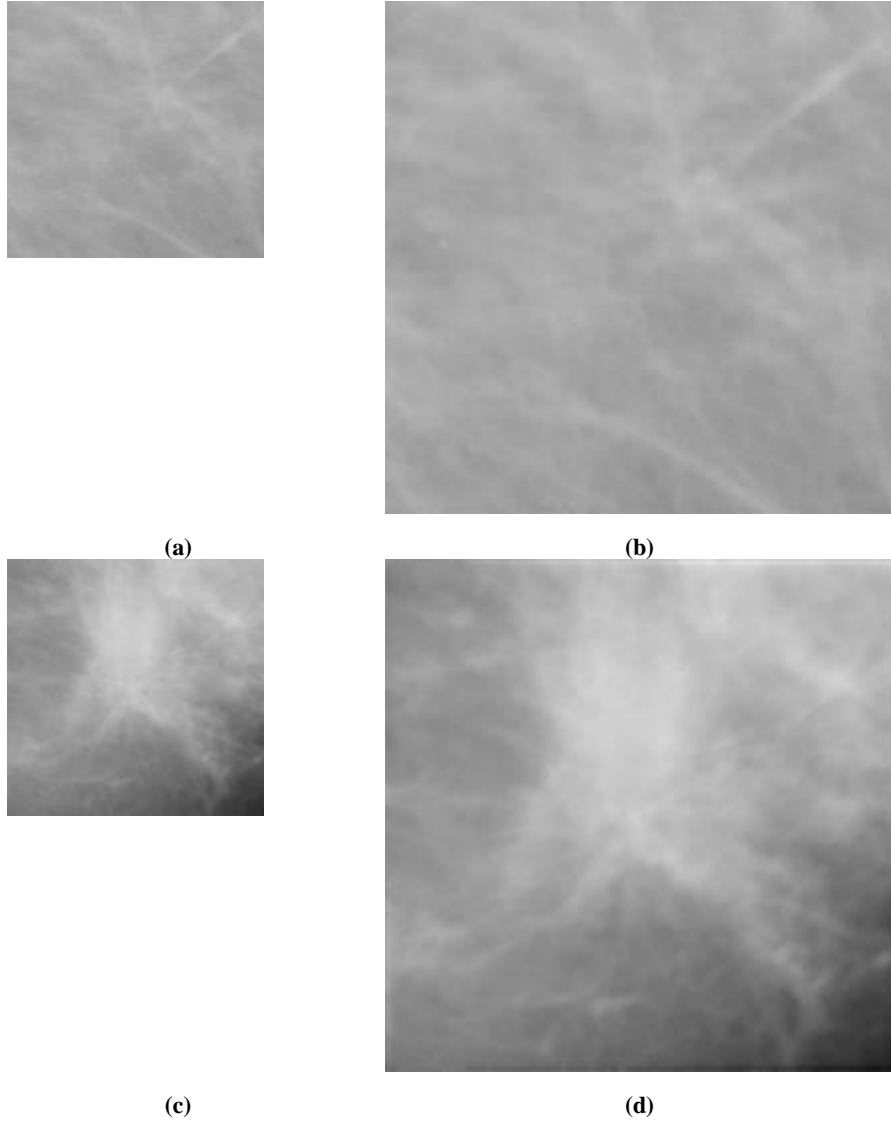


Figure A.2: Example2 for Mini-MIAS mammogram: (a) Normal LR-Mini-MIAS mammogram 128x128, (b) Normal SR-Mini-MIAS mammogram 256x256, (c) Abnormal LR-Mini-MIAS mammogram 128x128, and (d) Abnormal SR-Mini-MIAS mammogram 256x256

APPENDIX B: Super resolution Mini-MIAS database

Table B.1: Results of leave-one-out strategy based on LBPH features obtained from multi wavelet sub-bands

	LR-Mini-MIAS			SR-Mini-MIAS		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	94.69	99.04	97.52	93.81	98.56	96.89
LL,LH	96.46	99.04	98.14	94.69	99.52	97.83
LL,HL,LH	94.69	99.52	97.83	95.58	99.52	98.14
ALL	94.69	99.52	97.83	95.58	99.52	98.14

Table B.2: Results of 70% training and 30% testing strategy based on LBPH features obtained from multi wavelet sub-bands

	LR-Mini-MIAS			SR-Mini-MIAS		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	92.35	96.91	95.31	93.53	98.33	96.65
LL,LH	92.65	97.54	95.83	95.44	98.89	97.68
LL,HL,LH	92.65	98.41	96.39	95.00	99.05	97.63
ALL	94.12	99.37	97.53	94.12	99.13	97.37

Table B.3: Results of 50% training and 50% testing strategy based on LBPH features obtained from multi wavelet sub-bands

	LR-Mini-MIAS			SR-Mini-MIAS		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	89.38	96.39	93.94	94.02	97.88	96.53
LL,LH	90.00	97.02	94.56	93.93	98.27	96.75
LL,HL,LH	89.82	97.89	95.06	94.46	98.70	97.22
ALL	92.59	98.80	96.63	94.38	98.75	97.22

Table B.4: Results of 10% training and 90% testing strategy based on LBPH features obtained from multi wavelet sub-bands

LBP in Wavelet Sub-bands	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	81.23	94.92	90.10	89.90	97.37	94.74
LL,LH	81.32	94.97	90.17	90.74	96.91	94.74
LL,HL,LH	81.91	96.04	91.07	91.18	98.01	95.60
ALL	84.36	97.34	92.78	90.83	97.87	95.40

Table B.5: Results of different strategy based on LBPH features obtained from original mammogram mammograms

	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	92.92	96.65	95.34	96.46	97.61	97.20
70%/30% training/testing	91.07	95.69	94.08	93.97	97.38	96.19
50%/50% training/testing	90.18	95.22	93.46	95.96	96.14	96.08
10%/90% training/testing	88.24	94.15	92.07	91.23	95.98	94.31

Table B.6: Results of using four strategies based on HOG features obtained from original mammograms

	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	80.53	98.09	91.93	87.61	97.61	94.10
70%/30% training/testing	80.00	98.02	91.70	86.62	97.62	93.76
50%/50% training/testing	77.95	98.17	91.09	84.91	97.76	93.24
10%/90% training/testing	56.52	98.59	83.79	68.53	98.59	88.02

Table B.7: Results of using four strategies based on PCA-HOG and PCA-LBP features obtained from original mammogram

		LR-Mini-MIAS			SR-Mini-MIAS		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
PCA-HOG	Leave-one-Out	88.50	97.13	94.10	94.69	97.13	96.27
	70%/30% training/testing	88.38	96.83	93.87	94.41	96.03	95.46
	50%/ 50% training/testing	87.68	95.24	92.59	93.07	94.86	94.23
	10%/90% training/testing	73.53	95.27	87.62	81.18	91.73	88.02
PCA-LBP	Leave-one-Out	90.27	99.52	96.27	95.58	97.13	96.58
	70%/30% training/testing	88.82	99.52	95.77	94.85	97.46	96.55
	50%/50% training/testing	88.48	99.52	95.66	93.86	97.43	96.17
	10%/90% training/testing	68.19	100.00	88.81	85.88	96.01	92.45

Table B.8: Results of leave-one-out tests strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	97.35	99.04	98.45	97.35	97.61	97.52
O,HL	96.46	99.04	98.14	95.58	97.61	96.89
O,LH	95.58	99.04	97.83	97.35	95.69	96.27
O,LL,HL	97.35	99.04	98.45	97.35	98.09	97.83
O,LL,LH	97.35	99.04	98.45	96.46	97.61	97.20
O,LL,HL,LH	97.35	99.04	98.45	97.35	97.13	97.20
O,ALL	97.35	99.52	98.76	96.46	97.61	97.20

Table B.9: Results of 70% training and 30% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	96.77	99.21	98.35	94.56	98.41	97.06
O,HL	95.74	99.13	97.94	95.44	98.73	97.58
O,LH	95.59	98.97	97.78	94.26	98.33	96.91
O,LL,HL	96.47	99.21	98.25	94.26	98.25	96.86
O,LL,LH	96.47	99.21	98.25	95.00	98.49	97.27
O,LL,HL,LH	96.47	99.21	98.25	94.71	97.70	96.65
O,ALL	96.32	99.21	98.20	93.09	97.38	95.88

Table B.10: Results of 50% training and 50% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	95.36	98.80	97.59	94.73	96.68	96.00
O,HL	95.00	98.75	97.44	95.36	97.93	97.03
O,LH	95.09	98.75	97.47	94.38	97.36	96.31
O,LL,HL	95.45	98.94	97.72	95.71	97.93	97.16
O,LL,LH	95.27	98.99	97.69	94.20	96.01	95.38
O,LL,HL,LH	95.27	98.99	97.69	92.05	96.59	95.00
O,ALL	95.27	98.99	97.69	94.47	93.86	94.07

Table B.11: Results of 10% training and 90% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	91.28	98.88	96.21	90.93	96.22	94.36
O,HL	91.13	98.83	96.12	91.47	95.74	94.24
O,LH	91.18	98.88	96.17	90.34	96.25	94.17
O,LL,HL	91.32	98.88	96.22	91.23	95.29	93.86
O,LL,LH	91.28	98.88	96.21	90.29	96.01	94.00
O,LL,HL,LH	91.28	98.91	96.22	90.69	95.88	94.05
O,ALL	91.47	98.99	96.35	89.90	96.20	93.98

Table B.12: Results of using four strategies based on 14 GLCM features obtained from original mammogram

	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	96.46	100.00	98.76	86.73	95.22	92.24
70%/30% training/testing	86.03	98.10	93.87	85.59	92.06	89.79
50%/50% training/testing	91.70	98.46	96.09	91.25	89.95	90.41
10%/90% training/testing	82.26	71.30	75.16	57.30	66.62	63.34

Table B.13: Results of leave-one-out tests strategy based on decision fusing of using different kind of features

Fusions	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	96.46	99.52	98.45	97.35	99.52	98.76
<i>f2</i>	92.04	98.57	96.27	94.69	98.56	97.20
<i>f3</i>	98.23	99.52	99.07	97.35	99.52	98.76
<i>f4</i>	97.35	99.52	98.76	97.35	99.52	98.76
<i>f5</i>	92.92	98.09	96.27	94.69	97.13	96.27
<i>f6</i>	92.04	98.09	95.96	94.69	97.13	96.27
<i>f7</i>	96.46	99.52	98.45	96.46	99.04	98.14

Table B.14: Results of 70% training and 30% testing strategy based on decision fusing of using different kind of features

Fusions	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	93.97	99.37	97.47	96.91	98.73	98.09
<i>f2</i>	91.18	98.81	96.13	94.26	98.17	96.80
<i>f3</i>	95.15	99.29	97.84	97.94	99.05	98.66
<i>f4</i>	95.59	99.29	97.99	97.94	99.13	98.71
<i>f5</i>	92.65	98.73	96.60	95.00	96.19	95.77
<i>f6</i>	91.62	98.89	96.34	95.00	96.11	95.72
<i>f7</i>	92.79	99.52	97.17	97.06	98.73	98.14

Table B.15: Results of 50% training and 50% testing strategy based on decision fusing of using different kind of features

Fusions	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	94.12	99.38	97.53	97.37	98.33	97.99
<i>f2</i>	90.79	98.76	95.96	92.72	97.95	96.11
<i>f3</i>	95.53	99.33	97.99	98.60	97.19	97.69
<i>f4</i>	95.88	99.48	98.21	98.60	97.10	97.62
<i>f5</i>	92.37	98.19	96.14	93.33	95.10	94.48
<i>f6</i>	90.97	98.33	95.74	93.33	95.10	94.48
<i>f7</i>	92.63	99.62	97.16	93.26	95.96	95.01

Table B.16: Results of 10% training and 90% testing strategy based on decision fusing of using different kind of features

Fusions	LR-Mini-MIAS			SR-Mini-MIAS		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	84.27	99.55	94.17	81.67	98.09	92.31
<i>f2</i>	77.70	98.62	91.26	82.60	97.66	92.36
<i>f3</i>	86.28	99.47	94.83	86.76	94.81	91.98
<i>f4</i>	89.61	99.50	96.02	86.81	94.84	92.02
<i>f5</i>	80.15	97.82	91.60	81.27	91.78	88.09
<i>f6</i>	74.02	97.98	89.55	81.27	91.78	88.09
<i>f7</i>	74.61	99.84	90.97	86.47	94.73	91.83

APPENDIX C: Super resolution DDSM database

Table C.1: leave-one-out strategy based on LBPH features obtained from multi wavelet sub- bands

	LR-DDSM			SR-DDSM		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	87.50	89.06	88.28	95.31	96.88	96.09
LL,LH	88.67	90.63	89.65	95.31	98.05	96.68
LL,HL,LH	88.28	89.84	89.06	96.48	98.05	97.27
ALL	88.67	91.02	89.84	97.66	96.48	97.07

Table C.2: Results of 70% training and 30% testing strategy based on LBPH features obtained from multi wavelet sub-bands

	LR-DDSM			SR-DDSM		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	88.77	88.70	88.73	93.18	94.22	93.70
LL,LH	89.29	89.81	89.55	94.48	94.94	94.71
LL,HL,LH	89.16	89.87	89.51	97.01	95.58	96.30
ALL	88.57	90.33	89.45	95.39	95.58	95.49

Table C.3: Results of 50% training and 50% testing strategy based on LBPH features obtained from multi wavelet sub-bands

	LR-DDSM			SR-DDSM		
LBP in Wavelet Sub-bands	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	86.84	88.87	87.85	91.52	92.30	91.91
LL,LH	87.46	90.12	88.79	92.93	93.79	93.36
LL,HL,LH	86.56	89.65	88.11	94.53	94.96	94.75
ALL	86.88	89.65	88.26	95.43	93.98	94.71

Table C.4: Results of 10% training and 90% testing strategy based on LBPH features obtained from multi wavelet sub-bands

LBP in Wavelet Sub-bands	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LL,HL	83.72	86.00	84.86	84.91	87.78	86.35
LL,LH	85.33	86.17	85.75	85.50	88.59	87.04
LL,HL,LH	85.24	86.26	85.75	86.83	89.22	88.02
ALL	85.15	86.52	85.84	87.30	89.93	88.62

Table C.5: Results of different strategy based on LBPH features obtained from original mammogram

	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	88.67	92.58	90.63	98.44	93.36	95.90
70%/30% Training/testing	90.07	91.43	90.75	97.01	91.69	94.35
50%/50% training/testing	88.98	90.51	89.75	94.65	92.50	93.57
10%/90% training/testing	82.72	86.76	84.74	87.00	87.30	87.15

Table C.6: Results of using four strategies based on HOG features obtained from original mammogram

	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	72.66	77.34	75.00	77.73	81.25	79.49
70%/30% training/testing	71.56	78.57	75.07	73.57	80.97	77.27
50%/50% training/testing	69.49	80.00	74.75	70.90	82.03	76.46
10%/90% training/testing	60.13	77.85	68.99	59.61	71.41	65.51

Table C.7: Results of using four strategies based on PCA-HOG and PCA-LBP features obtained from original mammogram

		LR-DDSM			SR-DDSM		
		Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
PCA-HOG	Leave-one-Out	78.52	85.16	81.84	78.91	85.16	82.03
	70%/30% training/testing	75.58	84.22	79.90	77.60	83.51	80.55
	50%/ 50% training/testing	71.64	83.24	77.44	75.08	83.28	79.18
	10%/90% training/testing	63.09	74.46	68.77	67.57	73.96	70.76
PCA-LBP	Leave-one-Out	71.09	94.92	83.01	80.47	94.53	87.50
	70%/30% training/testing	68.05	93.44	80.75	80.00	92.60	86.30
	50%/50% training/testing	64.14	93.32	78.73	76.91	91.91	84.41
	10%/90% training/testing	40.98	92.78	66.88	56.59	89.65	73.12

Table C.8: Results of leave-one-out tests strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	90.63	92.58	91.60	98.44	94.14	96.29
O,HL	90.23	91.41	90.82	99.22	92.97	96.09
O,LH	91.02	92.19	91.60	98.44	91.02	94.73
O,LL,HL	90.23	92.58	91.41	98.44	92.97	95.70
O,LL,LH	90.23	93.36	91.80	97.27	94.53	95.90
O,LL,HL,LH	89.84	93.75	91.80	96.09	93.75	94.92
O,ALL	89.84	93.75	91.80	95.70	94.53	95.12

Table C.9: Results of 70% training and 30% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	89.74	92.53	91.14	96.43	93.77	95.10
O,HL	89.29	91.17	90.23	97.14	92.86	95.00
O,LH	89.55	91.69	90.62	96.49	89.94	93.21
O,LL,HL	89.74	92.66	91.20	97.21	90.71	93.96
O,LL,LH	90.00	92.86	91.43	98.12	88.31	93.21
O,LL,HL,LH	90.07	92.53	91.30	97.92	79.42	88.67
O,ALL	90.33	92.47	91.40	98.77	71.23	85.00

Table C.10: Results of 50% training and 50% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	89.92	90.66	90.29	94.69	94.41	94.55
O,HL	89.02	90.90	89.96	95.16	92.34	93.75
O,LH	89.10	91.21	90.16	94.84	91.84	93.34
O,LL,HL	89.73	90.63	90.18	96.72	89.49	93.11
O,LL,LH	89.77	90.78	90.27	94.88	93.05	93.96
O,LL,HL,LH	89.65	90.86	90.25	97.27	91.09	94.18
O,ALL	89.45	91.13	90.29	96.91	79.18	88.05

Table C.11: Results of 10% training and 90% testing strategy based on LBPH features obtained from original mammogram and wavelet sub-bands

LBP in Original and Wavelet Sub-bands	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
O,LL	84.00	87.48	85.74	88.57	85.17	86.87
O,HL	82.83	87.02	84.92	88.54	85.20	86.87
O,LH	83.13	86.87	85.00	88.07	85.80	86.93
O,LL,HL	84.04	87.52	85.78	87.48	86.24	86.86
O,LL,LH	84.24	87.50	85.87	88.09	86.61	87.35
O,LL,HL,LH	83.04	87.20	85.12	86.74	88.28	87.51
O,ALL	84.35	87.65	86.00	87.83	85.09	86.46

Table C.12: Results of using four strategies based on GLCM features obtained from original mammogram

	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Leave-one-Out	97.27	91.02	94.14	98.83	92.97	95.90
70%/30% training/testing	93.29	90.07	91.68	99.87	97.79	98.83
50%/50% training/testing	92.27	90.90	91.58	98.48	93.09	95.78
10%/90% training/testing	82.50	69.39	75.95	83.52	61.70	72.61

Table C.16: Results of leave-one-out tests strategy based on decision fusing of using different kind of features

Fusions	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	94.14	94.92	94.53	99.22	96.48	97.85
<i>f2</i>	83.59	91.41	87.50	89.06	93.36	91.21
<i>f3</i>	94.53	96.48	95.51	98.44	97.27	97.85
<i>f4</i>	94.14	96.48	95.31	98.05	98.83	98.44
<i>f5</i>	80.08	91.41	85.74	87.11	92.58	89.84
<i>f6</i>	78.13	90.63	84.38	84.38	92.97	88.67
<i>f7</i>	87.89	95.31	91.60	92.58	97.66	95.12

Table C.17: Results of 70% training and 30% testing strategy based on decision fusing of using different kind of features

Fusions	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	94.48	97.08	95.78	98.83	97.47	98.15
<i>f2</i>	82.40	91.30	86.85	87.14	92.34	89.74
<i>f3</i>	94.55	97.53	96.04	99.48	94.68	97.08
<i>f4</i>	95.00	97.40	96.20	99.16	95.45	97.31
<i>f5</i>	76.30	90.33	83.31	86.30	89.94	88.12
<i>f6</i>	75.65	90.33	82.99	83.70	91.43	87.56
<i>f7</i>	85.71	97.27	91.49	93.18	98.57	95.88

Table C.18: Results of 50% training and 50% testing strategy based on decision fusing of using different kind of features

Fusions	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	93.87	96.33	95.10	96.68	96.41	96.54
<i>f2</i>	79.45	91.37	85.41	84.61	92.34	88.48
<i>f3</i>	94.14	96.52	95.33	98.36	93.83	96.09
<i>f4</i>	93.95	96.41	95.18	97.93	94.45	96.19
<i>f5</i>	73.95	87.97	80.96	83.71	89.34	86.52
<i>f6</i>	72.85	87.73	80.29	80.74	90.51	85.63
<i>f7</i>	83.28	96.68	89.98	90.86	97.15	94.00

Table C.19: Results of 10% training and 90% testing strategy based on decision fusing of using different kind of features

Fusions	LR-DDSM			SR-DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	81.87	86.17	84.02	84.80	81.59	83.20
<i>f2</i>	70.63	85.59	78.11	75.22	83.20	79.21
<i>f3</i>	85.26	86.37	85.82	87.89	82.74	85.32
<i>f4</i>	85.65	86.83	86.24	89.00	83.80	86.40
<i>f5</i>	69.76	81.74	75.75	76.02	82.85	79.43
<i>f6</i>	67.09	82.15	74.62	70.20	83.83	77.01
<i>f7</i>	66.59	89.20	77.89	73.63	84.72	79.17

APPENDIX D: Feature Fusion

Table D.1: Results of 50% training and 50% testing strategy based on decision fusing of using different kind of features

Fusions	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	94.12	99.38	97.53	93.87	96.33	95.10
<i>f2</i>	90.79	98.76	95.96	79.45	91.37	85.41
<i>f3</i>	95.53	99.33	97.99	94.14	96.52	95.33
<i>f4</i>	95.88	99.48	98.21	93.95	96.41	95.18
<i>f5</i>	92.37	98.19	96.14	73.95	87.97	80.96
<i>f6</i>	90.97	98.33	95.74	72.85	87.73	80.29
<i>f7</i>	92.63	99.62	97.16	83.28	96.68	89.98

Table D.2: Results of 10% training and 90% testing strategy based on decision fusing of using different kind of features

Fusions	Mini-MIAS			DDSM		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
<i>f1</i>	84.27	99.55	94.17	81.87	86.17	84.02
<i>f2</i>	77.70	98.62	91.26	70.63	85.59	78.11
<i>f3</i>	86.28	99.47	94.83	85.26	86.37	85.82
<i>f4</i>	89.61	99.50	96.02	85.65	86.83	86.24
<i>f5</i>	80.15	97.82	91.60	69.76	81.74	75.75
<i>f6</i>	74.02	97.98	89.55	67.09	82.15	74.62
<i>f7</i>	74.61	99.84	90.97	66.59	89.20	77.89